

Open Research Online

The Open University's repository of research publications and other research outputs

Composition Sequences and Semigroups of Möbius Transformations

Thesis

How to cite:

Jacques, Matthew (2016). Composition Sequences and Semigroups of Möbius Transformations. PhD thesis The Open University.

For guidance on citations see [FAQs](#).

© 2016 The Author



<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Version: Version of Record

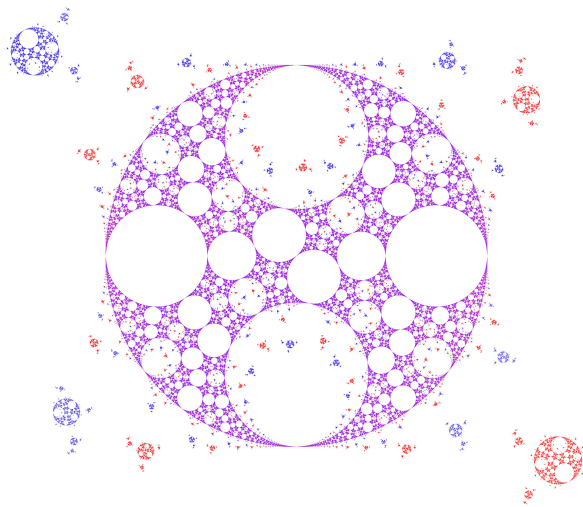
Link(s) to article on publisher's website:

<http://dx.doi.org/doi:10.21954/ou.ro.0000bd1f>

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

Composition sequences and semigroups of Möbius transformations



Matthew Jacques

MMath (Sheffield)

MSc (Open)

A thesis submitted to

SCHOOL OF MATHEMATICS AND STATISTICS

THE OPEN UNIVERSITY

for the degree of Doctor of Philosophy

January 23, 2017

Declaration

I confirm that the material contained in this thesis is the result of independent work, except where explicitly stated, and with the exception of Chapters 2, 3 and 5, which are the result of joint work with Ian Short. None of it has previously been submitted for a degree or other qualification to this or any other university or institution.

Matthew Jacques

January 23, 2017

Abstract

Motivated by the theory of Kleinian groups and by the theory of continued fractions, we study semigroups of Möbius transformations. Like Kleinian groups, semigroups have limit sets, and indeed each semigroup is equipped with two limit sets. We find that limit sets have an internal structure with features similar to the limit sets of Kleinian groups and the Julia sets of iterates of analytic functions. We introduce the notion of a *semidiscrete* semigroup, and find that this property is akin to the discreteness property for groups.

We study semigroups of Möbius transformations that fix the unit disc, and lay the foundations of a theory for such semigroups. We consider the composition sequences generated by such semigroups, and show that every such composition sequence converges pointwise in the open unit disc to a constant function whenever the identity element does not lie in the closure of the semigroup. We establish various results that have counterparts in the theory of Fuchsian groups. For example we show that aside from a certain exceptional family, any finitely-generated semigroup S is semidiscrete precisely when every two-generator semigroup contained in S is semidiscrete. We show that the limit sets of a nonelementary finitely-generated semidiscrete semigroup are equal (and non-trivial) precisely when the semigroup is a group. We classify two-generator semidiscrete semigroups, and give the basis for an algorithm that decides whether any two-generator semigroup is semidiscrete.

We go on to study finitely-generated semigroups of Möbius transformations that map the unit disc strictly within itself. Every composition sequence generated by such a semigroup converges pointwise in the open unit disc to a constant function. We give conditions that determine whether this convergence is uniform on the closed unit disc, and show that the cases where convergence is not uniform are very special indeed.

Acknowledgements

Firstly I would like to thank my supervisor, Ian Short, for his extensive help, encouragement, and boundless patience. I am very grateful for his support, without which I would have made little progress – thank you! I would also like to thank Phil Rippon, my second supervisor, for his generous support and advice. Everyone in the mathematics department has been extremely welcoming, and so I thank them, most especially everyone in the complex analysis group. I give particular thanks to the other Open University PhD students for their friendship and support, especially David Bevan, David Martí-Pete, Grahame, Mairi, Rosie and Vasso.

I would like to thank both examiners, Edward Crane (external) and David Brannan (internal), whose careful and extensive feedback on a previous version of this thesis has led to an immensely improved final version.

I also give a huge thank-you to my mother and brother for their continuous support. Finally, I would like to thank my partner, Anne-Marie, who encouraged me back into maths and has endured much while I have pursued it. This thesis is dedicated to her.

Contents

Declaration	i
Abstract	iii
Acknowledgements	v
Chapter 1. Introduction and background	1
1. Structure of this thesis	1
2. Continued fractions and composition sequences	2
3. The Möbius group and hyperbolic geometry	4
4. Sequences of Möbius transformations	9
5. Semigroups and Limit sets	10
Chapter 2. On Möbius semigroups	19
1. The group and inverse free parts of a semigroup	19
2. Semidiscrete semigroups	20
3. Schottky semigroups	21
4. Composition sequences	24
5. Covering regions	26
6. Constructing new semidiscrete semigroups from old	35
Chapter 3. Semigroups that fix the unit circle	39
1. A closer look at elementary subsemigroups of $\text{Aut}(\mathbb{H}^2)$	43
2. Proof of Theorem 3.1	46
3. Two-generator semigroups	49
4. Proof of Theorem 3.2	70
5. Classification of finitely-generated semigroups and proof of Theorem 3.3	78

6. Intersecting limit sets	83
Chapter 4. Limit sets	87
1. Introduction	87
2. On the structure of the limit sets	88
3. On semigroups with disjoint limit sets	105
Chapter 5. Self-maps of the disc	123
1. Introduction	123
2. Eventually tangent sequences of discs	129
3. Necessary and sufficient conditions for limit-disc type	130
4. Hausdorff dimension of the set of composition sequences of limit-disc type	133
5. Complete tangency graph	133
6. Convergence of composition sequences	141
Chapter 6. Coding limit sets	145
Bibliography	159

CHAPTER 1

Introduction and background

1. Structure of this thesis

In this first chapter we introduce the preliminary tools and ideas used in subsequent chapters, as well as giving some historical background. We shall begin by giving the classical definition of a continued fraction, which we shall later regard as a composition sequence of Möbius transformations, before giving a brief overview of hyperbolic geometry. We then introduce semigroups of Möbius transformations and discuss recent work of Fried, Marotta and Stankewitz [15].

Chapter 2 builds on the theory of semigroups of Möbius transformations presented in [15]. In Chapter 3 we study semidiscrete semigroups of Möbius transformations that fix the unit disc. These can be regarded as semigroup analogues of Fuchsian groups. The material in Chapters 2 and 3 is used in a joint paper [21] with Ian Short. Theorem 2.18 in Chapter 2 is due to Edward Crane. In Chapter 4 we study the limit sets of semigroups in detail. In Chapter 5 we turn to composition sequences generated by finitely many Möbius transformations that each map the unit disc strictly within itself. The material in this chapter is used in a joint paper [22] with Ian Short. Finally in Chapter 6 we consider what can be regarded as a generalisation of the continued fraction expansion of a real number. The work contained in Chapter 6 is in progress.

Throughout we shall freely assume the axiom of choice, which is used in, for example, König's lemma and the Baire category theorem.

2. Continued fractions and composition sequences

We denote the positive integers by \mathbb{N} . A *continued fraction* is a formal expression

$$\mathbf{K}(a_n|b_n) = \frac{a_1}{b_1 + \frac{a_2}{b_2 + \cdots}},$$

where a_n and b_n are infinite sequences of complex numbers and $a_n \neq 0$ for each $n \in \mathbb{N}$.

We say that the continued fraction *converges* if the sequence with n^{th} term

$$\frac{a_1}{b_1 + \frac{a_2}{\ddots + \frac{a_n}{b_n}}}$$

converges in the extended complex plane $\overline{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$. We shall always consider continued fractions as generated by infinite (rather than finite) sequences a_n and b_n ; in particular we are excluding terminating continued fractions.

For a general choice of a_n and b_n the continued fraction may not converge. It is well known that if $a_n = 1$ and $b_n \in \mathbb{N}$ for all n , then the continued fraction does converge. Indeed, there is a one-to-one correspondence between each sequence b_1, b_2, \dots in \mathbb{N} and the irrational numbers contained in $(0, 1)$ via the function that maps b_n to the limit of the sequence

$$\frac{1}{b_1 + \frac{1}{\ddots + \frac{1}{b_n}}}$$

We equip the open unit ball $\mathbb{B}^3 = \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 < 1\}$ with the hyperbolic metric, denoted ρ , which is induced by the line element

$$ds^2 = \frac{4|d\mathbf{x}|^2}{(1 - |\mathbf{x}|^2)^2},$$

where \mathbf{x} is a point in \mathbb{R}^3 . This means that, for points \mathbf{x} and \mathbf{y} in \mathbb{B}^3 , the distance $\rho(\mathbf{x}, \mathbf{y})$ is given by the infimum of the integral of this metric along all curves between \mathbf{x} and \mathbf{y} .

With respect to ρ there is a unique geodesic between any two points in \mathbb{B}^3 . Every such geodesic is an arc of a Euclidean circle orthogonal to the unit sphere, \mathbb{S}^2 .

We define the ideal boundary of (\mathbb{B}^3, ρ) as follows. A *geodesic half-ray* is a map $\gamma : [0, \infty) \rightarrow \mathbb{B}^3$ such that $\rho(\gamma(0), \gamma(t)) = t$ for all $t \in [0, \infty)$. We define a relation \sim on the set of geodesic half-rays as follows. If γ_1 and γ_2 are geodesic half-rays, then $\gamma_1 \sim \gamma_2$ if there is a constant $C > 0$ such that $\rho(\gamma_1(t), \gamma_2(t)) < C$ for all t . It is easy to check that \sim is an equivalence relation. The *ideal boundary* of (\mathbb{B}^3, ρ) is defined to be the set of equivalence classes of geodesic half-rays, and can be identified with the unit sphere \mathbb{S}^2 .

We consider the group $\text{Aut}(\mathbb{B}^3)$ of Möbius transformations, defined to be the orientation-preserving isometries of (\mathbb{B}^3, ρ) . The action of $\text{Aut}(\mathbb{B}^3)$ on \mathbb{B}^3 extends to a conformal action on its boundary \mathbb{S}^2 , which, by stereographic projection, we identify with the extended complex plane $\overline{\mathbb{C}}$, as we explain shortly. It is well known that $\text{Aut}(\mathbb{B}^3)$ is exactly the group of conformal automorphisms of $\overline{\mathbb{C}}$. Moreover, when regarded as functions acting on the complex plane, these Möbius transformations are exactly those maps of the form

$$z \mapsto \frac{az + b}{cz + d},$$

where $a, b, c, d \in \mathbb{C}$ and $ad - bc = 1$.

Given a collection \mathcal{F} of Möbius transformations, we define a *composition sequence generated by \mathcal{F}* to be a sequence

$$F_n = f_1 \cdots f_n,$$

where each $f_i \in \mathcal{F}$. If $f_n(z) = \frac{a_n}{z + b_n}$ for each n , then the continued fraction $\mathbf{K}(a_n|b_n)$ converges exactly when $F_n(0)$ converges. This is the connection between composition sequences of Möbius transformations and continued fractions. In this thesis we use hyperbolic geometry, and other methods, to study composition sequences in detail. We are particularly interested in the relationship between the collection of composition sequences generated by some subset \mathcal{F} of $\text{Aut}(\mathbb{B}^3)$ and the semigroup generated by \mathcal{F} .

Regarding continued fractions as composition sequences of Möbius transformations is useful in at least two ways. Firstly, the tools of hyperbolic geometry can be applied directly to the study of continued fractions, which can allow for the replacement of opaque algebraic arguments with more concise geometric alternatives. Secondly, composition sequences give a natural way of generalising continued fractions. For example, one can also consider composition sequences of Möbius transformations in higher dimensions. In the sequel we work with Möbius transformations acting on \mathbb{B}^3 and \mathbb{S}^2 , although many of our results also hold for Möbius transformations acting on \mathbb{B}^n and \mathbb{S}^{n-1} . Following Aebischer's 1990 paper [1], several authors, notably Beardon, Hockman and Short (see for example [4]), have engaged in a programme of interpreting and extending classical results from continued fraction theory using the tools of hyperbolic geometry.

3. The Möbius group and hyperbolic geometry

We have already introduced $\text{Aut}(\mathbb{B}^3)$ as the group of orientation-preserving isometries of (\mathbb{B}^3, ρ) ; however, we now describe a more general collection of Möbius transformations. We let $\overline{\mathbb{R}^3} = \mathbb{R}^3 \cup \{\infty\}$ be the one-point compactification of \mathbb{R}^3 . We denote the group of conformal homeomorphisms of $\overline{\mathbb{R}^3}$ by $\text{Aut}(\overline{\mathbb{R}^3})$. The group $\text{Aut}(\overline{\mathbb{R}^3})$ is the set of orientation-preserving maps in the group that is generated by reflections in planes and inversions in spheres acting on $\overline{\mathbb{R}^3}$. The set $\text{Aut}(\mathbb{B}^3)$ is the subgroup of elements in $\text{Aut}(\overline{\mathbb{R}^3})$ that fix \mathbb{B}^3 setwise. We denote the distinguished point $(0, 0, 0)$ in the unit ball by $\mathbf{0}$.

Occasionally it will be convenient to use a different model of hyperbolic space. We define *upper half-space* by $\mathbb{H}^3 = \{(x, y, t) \mid (x, y) \in \mathbb{R}^2, t > 0\}$ and denote the distinguished point $(0, 0, 1) \in \mathbb{H}^3$ by \mathbf{j} . The hyperbolic metric on \mathbb{H}^3 , which we shall also denote by ρ , is induced by the line element

$$ds^2 = \frac{|d\mathbf{x}|^2}{t^2},$$

where \mathbf{x} is a point in \mathbb{H}^3 . Equipped with this metric, we call \mathbb{H}^3 the *upper half-space model*. Exploiting the same construction used to define the ideal boundary of (\mathbb{B}^3, ρ) , we find that (\mathbb{H}^3, ρ) has a well-defined ideal boundary that we identify with $\overline{\mathbb{C}}$. Choose an

element $\phi \in \text{Aut}(\overline{\mathbb{R}^3})$ that maps \mathbb{B}^3 onto \mathbb{H}^3 , and maps the points $\mathbf{0}$, $(0, 0, 1)$ and $(0, 0, -1)$ to \mathbf{j} , ∞ and 0 respectively. Upon conjugating $\text{Aut}(\mathbb{B}^3)$ by ϕ , we obtain the subgroup of $\text{Aut}(\overline{\mathbb{R}^3})$ that fixes \mathbb{H}^3 . This is exactly the group of orientation-preserving isometries of (\mathbb{H}^3, ρ) .

We shall routinely switch between the upper half-space and unit ball models of hyperbolic space, and do not make a notational distinction between a Möbius transformation acting on one model and its corresponding conjugate acting on another. The upper half-space model is usually preferred in an argument that considers a distinguished point on the ideal boundary, which by conjugation we can take to be ∞ . On the other hand if our argument involves a distinguished point in hyperbolic space, the unit ball model is useful so that, by conjugation, this distinguished point may be taken to be $\mathbf{0}$.

The above discussion is dimensionless, although the number of conditions needed to identify a unique Möbius transformation from \mathbb{H}^n onto \mathbb{B}^n depends on n . In two dimensions, we have the unit disc instead of the unit ball, which, when thought of as a subset of the complex plane, carries the hyperbolic metric (also denoted ρ) given by the line element

$$ds^2 = \frac{4|dz|^2}{(1 - |z|^2)^2},$$

where $z \in \mathbb{D}$. We call (\mathbb{D}, ρ) the *unit disc model* of the hyperbolic plane. The group of Möbius transformations acting on $\overline{\mathbb{C}}$ that fix the unit disc is denoted $\text{Aut}(\mathbb{D})$, and is exactly the group of orientation-preserving isometries of (\mathbb{D}, ρ) . Choose a Möbius transformation ϕ that maps the unit disc onto the upper half-plane $\mathbb{H}^2 = \{(x, y) \mid y > 0\}$, and maps 0 to i . The map ϕ transfers the hyperbolic metric from the unit disc onto the upper half-plane. In this model the hyperbolic metric is given by the line element

$$ds^2 = \frac{|dz|^2}{\text{Im}(z)^2},$$

where $z \in \mathbb{H}^2$. By conjugating $\text{Aut}(\mathbb{D})$ by ϕ we obtain the collection of orientation-preserving isometries of the upper half-plane.

Many of our results hold for Möbius transformations acting in higher dimensions. However, for clarity of exposition we shall always work with Möbius transformations acting on at most three dimensions. If however a result is stated to hold for Möbius transformations belonging to $\text{Aut}(\mathbb{D})$ (or $\text{Aut}(\mathbb{H}^2)$), then we only claim the proof holds for these cases. Many results in Chapter 3 are of this type.

We let χ denote the chordal metric on the closed unit ball, which is defined as the restriction of the Euclidean metric on \mathbb{R}^3 to the closed unit ball. The unit sphere \mathbb{S}^2 can be identified with $\overline{\mathbb{C}}$ by the well-known *stereographic projection* map, which we now describe. A point $\mathbf{x} \in \mathbb{S}^2 \setminus \{(0, 0, 1)\}$ is first mapped to $(x, y, 0) \in \mathbb{R}^3$, defined as the point where the plane $\{(x, y, z) \in \mathbb{R}^3 \mid z = 0\}$ intersects the infinite line in \mathbb{R}^3 that passes through \mathbf{x} and $(0, 0, 1)$. The stereographic projection map sends $\mathbf{x} \in \mathbb{S}^2 \setminus \{(0, 0, 1)\}$ to $x + iy \in \overline{\mathbb{C}}$, while the point $(0, 0, 1)$ is mapped to $\infty \in \overline{\mathbb{C}}$. If z and w are points in $\overline{\mathbb{C}}$ then $\chi(z, w)$ is given by

$$(1) \quad \chi(z, w) = \frac{2|z - w|}{\sqrt{1 + |z|^2}\sqrt{1 + |w|^2}},$$

where it is understood that the appropriate limit is taken if one or both of z or w equals ∞ . We now describe a complete metric on the group $\text{Aut}(\mathbb{B}^3)$ itself. For $f, g \in \text{Aut}(\mathbb{B}^3)$, the *metric of uniform convergence* is given by

$$(2) \quad \sigma(f, g) = \sup_{z \in \mathbb{B}^3} |f(z) - g(z)|.$$

Directly from its definition, σ can be seen to be right invariant, that is

$$\sigma(fh, gh) = \sigma(f, g)$$

for all $f, g, h \in \text{Aut}(\mathbb{B}^3)$. In fact $\text{Aut}(\mathbb{B}^3)$ is a topological group [4, Theorem 3.1], that is, the maps $f \mapsto f^{-1}$ from $\text{Aut}(\mathbb{B}^3)$ to itself, and $(f, g) \mapsto fg$ from $\text{Aut}(\mathbb{B}^3) \times \text{Aut}(\mathbb{B}^3)$ to $\text{Aut}(\mathbb{B}^3)$ are both continuous.

The map Φ given by

$$\Phi \left(\begin{pmatrix} a & b \\ c & d \end{pmatrix} \right) = \left(z \mapsto \frac{az + b}{cz + d} \right)$$

from $\mathrm{SL}(2, \mathbb{C})$ onto $\mathrm{Aut}(\mathbb{B}^3)$ is a surjective group homomorphism with kernel $\{I, -I\}$, where I is the identity matrix. This implies that $\mathrm{Aut}(\mathbb{B}^3)$ is isomorphic to $\mathrm{PSL}(2, \mathbb{C}) = \mathrm{SL}(2, \mathbb{C}) / \{I, -I\}$.

For any Möbius transformation f the square of the trace of both representations of f in $\mathrm{SL}(2, \mathbb{C})$ is the same. This implies that the function $\mathrm{tr}^2 : \mathrm{Aut}(\mathbb{B}^3) \rightarrow \mathbb{C}$ given by $\mathrm{tr}^2(f) = (\mathrm{trace}(F))^2$, where $F \in \mathrm{SL}(2, \mathbb{C})$ satisfies $\Phi(F) = f$, is well defined. Indeed, Φ is also an open, continuous and surjective map [4, Theorem 3.2].

The function tr^2 is important because two elements $f, g \in \mathrm{Aut}(\mathbb{B}^3)$ are conjugate if and only if $\mathrm{tr}^2(f) = \mathrm{tr}^2(g)$ (see [4, Theorem 3.10]). We let I denote the identity map. A Möbius transformation f not equal to the identity falls into one of the following three classes:

- The map f is *parabolic* if f is conjugate to the map $z \mapsto z + 1$.
In this case $\mathrm{tr}^2(f) = 4$ and f has exactly one fixed point in $\overline{\mathbb{C}}$ with multiplier equal to 1.
- The map f is *loxodromic* if f is conjugate to a map of the form $z \mapsto \lambda z$ where $|\lambda| \neq 1$.
In this case $\mathrm{tr}^2(f) \in \mathbb{C} \setminus [0, 4]$ and f has exactly two fixed points in $\overline{\mathbb{C}}$, one attracting and one repelling.
- The map f is *elliptic* if f is conjugate to a map of the form $z \mapsto \lambda z$ where $|\lambda| = 1$ and $\lambda \neq 1$.
In this case $\mathrm{tr}^2(f) \in [0, 4)$ and f has exactly two fixed points in $\overline{\mathbb{C}}$, each with multiplier of modulus 1.

It is useful to observe that the function tr^2 is continuous. Since Φ is an open map, it follows that the collection of loxodromic elements is an open subset of $\mathrm{Aut}(\mathbb{B}^3)$.

If a Möbius transformation g satisfies $g^n = I$ for some positive integer n , then we call the least such n the *order* of g , and write $n = \mathrm{order}(g)$. If there is no such positive integer n , then we say that g has *infinite order*. By the characterisation of Möbius transformations given above, it is clear that only elliptic transformations conjugate to rational rotations

have finite order. All other nontrivial Möbius transformations have infinite order.

Let us now give a geometric description of the action of each type of Möbius transformation on hyperbolic space. Each elliptic and loxodromic transformation has two fixed points on the ideal boundary. The hyperbolic geodesic landing at these fixed points is called the *axis* of the transformation. Each elliptic transformation fixes its axis pointwise. The action of the elliptic transformation is to perform a nontrivial hyperbolic rotation around its axis. The angle of rotation, θ , of an elliptic transformation f is related to $\text{tr}^2(f)$ by the equation $\text{tr}^2(f) = 4\cos^2(\theta/2)$. A loxodromic transformation f has one attracting fixed point and one repelling fixed point, which we shall denote by α_f and β_f respectively. The map f first translates each point x in \mathbb{B}^3 by some nonzero hyperbolic distance, called the *translation length* of f , parallel to its axis and in the direction of the attracting fixed point of f , before performing a (possibly trivial) hyperbolic rotation about its axis. A loxodromic transformation fixes its axis setwise. The action of a parabolic transformation is very different. A *horoball* is a Euclidean sphere internally tangent (and not equal) to the unit sphere, and we call the point of tangency the *base point* of the horoball. A parabolic transformation fixes any horoball based at its fixed point setwise. The action of a parabolic element f is best visualised by thinking in the upper half-space model, and conjugating f to the map $z \mapsto z + 1$, which fixes the horoball $\{(x, y, t) \mid (x, y) \in \mathbb{R}^2 \cup \{\infty\}\}$ for any given $t > 0$.

To understand the action of a Möbius transformation on the ideal boundary, the notion of isometric discs can be helpful. First, we define the *chordal derivative* of a Möbius transformation f at $w \in \overline{\mathbb{C}}$ to be the quantity

$$\lim_{z \rightarrow w} \frac{\chi(f(z), f(w))}{\chi(z, w)}.$$

For any Möbius transformation f , its *isometric disc*, which we shall denote by $I^+(f)$, is the open chordal disc contained in $\overline{\mathbb{C}}$ where the chordal derivative of f is greater than 1. We let $I^-(f)$ denote the isometric disc of f^{-1} . If f is loxodromic then $\alpha_f \in I^-(f)$ and $\beta_f \in I^+(f)$. Both isometric discs have chordal radius $1/\sinh[\frac{1}{2}\rho(f(\mathbf{0}), \mathbf{0})]$. The transformation f maps the interior of the complement of $I^+(f)$ onto $I^-(f)$. See [29, Chapter 1] for more details on the action of Möbius transformations on hyperbolic space and the

properties of isometric discs.

Definition A *Kleinian group* is a discrete group of Möbius transformations. A *Fuchsian group* is a Kleinian group whose elements all fix some non-trivial chordal disc in $\overline{\mathbb{C}}$. Equivalently, a Fuchsian group is a Kleinian group that is conjugate to a discrete subgroup of $\text{Aut}(\mathbb{D})$ by a Möbius transformation.

We recall the following definition of a Schottky group, which may be found in [29, Section 2.7].

Definition Let $\{C_1, C'_1, \dots, C_k, C'_k\}$ be k pairs of circles in $\overline{\mathbb{C}}$, whose interiors are pairwise disjoint. Suppose that for each $n = 1, \dots, k$ the Möbius transformation g_n maps C_n onto C'_n , and maps the interior of C_n onto the exterior of C'_n . The group generated by $\{g_1, \dots, g_k\}$ is a discrete group. We say that a group of Möbius transformations is a *Schottky group* if it is conjugate to a group constructed as above.

For some authors the definition given above is that of a *classical* Schottky group, and they give a more general definition of a Schottky group, where the circles featured in the construction are replaced by Jordan curves.

Schottky groups are examples of *free groups*. A free group is a group that has no non-trivial relations. We shall also need the related but slightly different notion of a *free semigroup*. We shall say that a semigroup S generated by a set \mathcal{F} of Möbius transformations is free if each element in S can be written uniquely as a finite composition of elements in $\mathcal{F} \setminus \{I\}$.

4. Sequences of Möbius transformations

A sequence of Möbius transformations g_n is said to *converge ideally* if for any point $\zeta \in \mathbb{B}^3$ the sequence $g_n(\zeta)$ converges in the Euclidean metric to a point on the ideal boundary. This definition is independent of the choice of point ζ in hyperbolic space. We say ‘converges ideally’ rather than simply ‘converges’, as the latter term is reserved to describe the convergence of g_n as a sequence inside the Möbius group, that is, if g_n converges uniformly to a Möbius transformation. Ideal convergence was introduced to the theory of continued

fractions by Lorentzen (formerly known as Jacobson) in [26], although there the term *general convergence* was used, and it was defined purely in terms of the action of $\text{Aut}(\mathbb{B}^3)$ on $\overline{\mathbb{C}}$.

The notion of an ideally converging sequence of Möbius transformations will play an important role in this thesis. It is useful to define another, weaker property on sequences of Möbius transformations. We say that a sequence g_n in $\text{Aut}(\mathbb{B}^3)$ is *escaping* if for any $\zeta \in \mathbb{B}^3$ the sequence $g_n(\zeta)$ accumulates only on the boundary of hyperbolic space, with respect to the Euclidean metric on $\overline{\mathbb{B}^3}$. If a sequence g_n is escaping then we call the sequence an *escaping sequence*. Again, since each g_n is a hyperbolic isometry, this definition is independent of the choice of ζ . In the continued fractions literature, what we call escaping sequences are sometimes called ‘restrained sequences’. These were first defined (see [28, Definition 2.6]) by Lorentzen and Thron purely in terms of the action of a sequence on $\overline{\mathbb{C}}$.

5. Semigroups and Limit sets

We define a *semigroup of Möbius transformations*, or more briefly, a *semigroup*, to be a set of Möbius transformations that is closed under composition. Note that the identity element, which we shall denote by I , need not lie in a semigroup.

Throughout the thesis, we write $A \subseteq B$ to mean that A is contained in or equal to B . If A is contained in but not equal to B , then we shall write $A \subsetneq B$. For any collection $X \subseteq \text{Aut}(\mathbb{B}^3)$ of Möbius transformations and any $\zeta \in \overline{\mathbb{B}^3}$, we describe the set of points $X^{-1}(\zeta) = \{g^{-1}(\zeta) \mid g \in X\}$ as the *backward orbit* of ζ under X . Similarly, $X(\zeta) = \{g(\zeta) \mid g \in X\}$ is the *forward orbit* of ζ under X , or the *X -orbit* of ζ . We now introduce one of the central ideas within this thesis.

Definition Choose any point $\zeta \in \mathbb{B}^3$. We define the *backward limit set* of X as

$$\Lambda^-(X) = \overline{X^{-1}(\zeta)} \cap \mathbb{S}^2,$$

where the closure is with respect to the Euclidean metric. Similarly we define the *forward limit set* of X as

$$\Lambda^+(X) = \overline{X(\zeta)} \cap \mathbb{S}^2.$$

These definitions are independent of the choice of ζ in hyperbolic space. Both limit sets are closed subsets of \mathbb{S}^2 . If g_n is a sequence then we shall refer to $\Lambda^-(\{g_n \mid n \in \mathbb{N}\})$ as the backward limit set of g_n , and abbreviate this simply to $\Lambda^-(g_n)$. Similarly we write $\Lambda^+(g_n)$ for $\Lambda^+(\{g_n \mid n \in \mathbb{N}\})$ and call this the forward limit set of g_n . When X is a Kleinian group, the forward and backward limit sets coincide with the limit set of the Kleinian group. A good introduction to Kleinian groups and the properties of their limit sets can be found in [3] or [29]. We shall be interested in limit sets where X is either a semigroup, or a sequence of Möbius transformations. Indeed, the limit sets of a semigroup will be of great interest to us, and we shall study them in detail. Limit sets of semigroups were studied by Fried, Marotta and Stankewitz in [15], and we shall develop their ideas. Limit sets of sequences are important because of the following theorem due to Aebischer [1], which, for a sequence of Möbius transformations g_n , connects convergence of the sequence $g_n(z)$ where z is a point on the ideal boundary, to convergence of $g_n(\zeta)$ where ζ is a point in hyperbolic space. See also [4, Theorem 9.6].

THEOREM 1.1. *Suppose that g_n is an escaping sequence of Möbius transformations. Then $\chi(g_n(z), g_n(\mathbf{0})) \rightarrow 0$ as $n \rightarrow \infty$ for all $z \in \overline{\mathbb{B}^3} \setminus \Lambda^-(g_n)$. In fact convergence is locally uniform in $\overline{\mathbb{B}^3} \setminus \Lambda^-(g_n)$, and $\overline{\mathbb{B}^3} \setminus \Lambda^-(g_n)$ is the largest open set with this property.*

In fact, it is shown in [1] that for any X the complement of $\Lambda^-(X)$ is the largest open subset of the ideal boundary upon which X is a normal family.

We now define two important subsets of the backward and forward limit sets, which have familiar counterparts in the theory of Kleinian groups. Suppose that X is a set of Möbius transformations and let $z \in \mathbb{S}^2$. Choose any hyperbolic geodesic segment γ that has one end point z and the other end point in \mathbb{B}^3 , and let $\zeta \in \mathbb{B}^3$. Then z is a *forward conical limit point* of X if there is sequence g_n in X , and a positive constant M , such that

- (i) $g_n(\zeta) \rightarrow z$ as $n \rightarrow \infty$;
- (ii) $\rho(\gamma, g_n(\zeta)) < M$ for all n .

For such a sequence g_n , we shall say that $g_n(\zeta)$ *converges conically* to z . These definitions do not depend on the choice of γ or ζ . The *forward conical limit set* $\Lambda_c^+(X)$ of X is the collection of all forward conical limit points of X . A *backward conical limit point* of X is defined to be a forward conical limit point of X^{-1} , and the *backward conical limit set*

$\Lambda_c^-(X)$ of X is defined to be $\Lambda_c^+(X^{-1})$.

We also define the *forward horospherical limit set* $\Lambda_h^+(X)$ of X to consist of those points z in $\Lambda^+(X)$ such that, for any point ζ in \mathbb{B}^3 , the orbit $X(\zeta)$ meets every horoball based at z . The *backward horospherical limit set* $\Lambda_h^-(X)$ of X is defined as $\Lambda_h^+(X^{-1})$. Clearly we have the inclusions

$$\Lambda_c^+ \subseteq \Lambda_h^+ \subseteq \Lambda^+ \quad \text{and} \quad \Lambda_c^- \subseteq \Lambda_h^- \subseteq \Lambda^-.$$

Here we have omitted mention of X , as we often will do for brevity. Conical and horospherical limit sets have important roles in the theory of Kleinian groups, so it is no surprise that the sets introduced here play an important part in the study of semigroups.

Given a semigroup S , we say that a subset Y of $\overline{\mathbb{B}^3}$ is *forward invariant* (with respect to S) if for each element g of S , $g(Y) \subseteq Y$. Similarly we shall say that $Y \subseteq \overline{\mathbb{B}^3}$ is *backward invariant* if for each element g of S , $g^{-1}(Y) \subseteq Y$. Equivalently, Y is backward invariant with respect to S if Y is forward invariant with respect to S^{-1} . In [15] the authors make the important observation that $\Lambda^+(S)$ is forward invariant and $\Lambda^-(S)$ is backward invariant. Indeed, the conical and horospherical limit sets behave in the same way: $\Lambda_c^+(S)$ and $\Lambda_h^+(S)$ are forward invariant, while $\Lambda_c^-(S)$ and $\Lambda_h^-(S)$ are backward invariant, which one can verify directly from the definitions.

The backward conical limit set of sequences features in the theory of continued fractions because of the following theorem of Aebischer [1, Theorem 5.2].

THEOREM 1.2. *Let g_n be an escaping sequence, and let $\zeta \in \mathbb{B}^3$. Then, for each point z in \mathbb{S}^2 , we have $\chi(g_n(\zeta), g_n(z)) \rightarrow 0$ as $n \rightarrow \infty$ if and only if $z \notin \Lambda_c^-(g_n)$.*

It is straightforward to prove this theorem using hyperbolic geometry; see, for example, [11, Proposition 3.3]. We also have the following lemma, proven in [11, Lemma 3.4].

LEMMA 1.3. *Suppose that w is a backward conical limit point of an escaping sequence g_n . Then there is a sequence of positive integers n_1, n_2, \dots and two distinct points p and q in \mathbb{S}^2 such that $g_{n_i}(w) \rightarrow p$ and $g_{n_i}(z) \rightarrow q$ for every point z in \mathbb{S}^2 other than w .*

Clearly any group is a semigroup with equal backward and forward limit sets. In general the limit sets of a semigroup do not coincide, and we shall study how the two sets relate to each other and what they can tell us about the underlying semigroup.

For any set $\mathcal{F} \subseteq \text{Aut}(\mathbb{B}^3)$ we shall write $\langle \mathcal{F} \rangle$ to denote the semigroup consisting of all finite compositions of elements that belong to \mathcal{F} . We shall call $\langle \mathcal{F} \rangle$ the *semigroup generated by* \mathcal{F} . If $S = \langle \mathcal{F} \rangle$ for some finite collection \mathcal{F} of Möbius transformations, then we say that S is *finitely-generated*. To be concise we denote the semigroup generated by elements $f, g \in \text{Aut}(\mathbb{B}^3)$ by $\langle f, g \rangle$, rather than by $\langle \{f, g\} \rangle$.

It is often convenient to conjugate a semigroup by a Möbius transformation. One reason this is useful is because, as is well known, the group of Möbius transformations is transitive on distinct ordered triples. That is, for any two ordered lists of three distinct points in $\overline{\mathbb{C}}$, (p, q, r) and (p', q', r') say, there is a unique Möbius transformation that maps p, q and r to p', q' and r' respectively. If S is a semigroup then we can conjugate S by any $f \in \text{Aut}(\mathbb{B}^3)$ to give another semigroup

$$T = fSf^{-1} = \{fgf^{-1} \mid g \in S\}.$$

It is clear that $\Lambda^+(T) = f(\Lambda^+(S))$ and $\Lambda^-(T) = f(\Lambda^-(S))$. Throughout this thesis whenever we say that a semigroup S is conjugate to a semigroup T we shall mean conjugate by an element in $\text{Aut}(\mathbb{B}^3)$. If S and T are semigroups and $S \subseteq T$, then we shall describe S as a *subsemigroup* of T . Finally we observe that if S is a semigroup, then so is S^{-1} .

In [15], Fried, Marotta and Stankewitz introduce semigroups of Möbius transformations and backward and forward limit sets. They go on to show that if the backward and forward limit sets of a semigroup are disjoint, then the semigroup can be regarded as an iterated function system on a certain metric space, which they construct. Recall that the limit set $\Lambda(G)$ of a Kleinian group G is either a perfect set, or contains 0, 1 or 2 points. Moreover, a group G is said to be elementary if $\Lambda(G)$ is not a perfect set, or equivalently, if there exists a point $x \in \overline{\mathbb{B}^3}$ such that the G -orbit of x is finite. The following theorem, given in

[15, Theorem 2.11] shows that just as for groups, semigroups of Möbius transformations come in varieties that can be regarded as elementary and nonelementary.

THEOREM 1.4. *Let S be a semigroup. Then either $\Lambda^+(S)$ is a perfect set or S is conjugate in $\text{Aut}(\mathbb{B}^3)$ to a semigroup T of one of the following types.*

- (i) $|\Lambda^+(T)| = 0$ and T is a semigroup of elliptic maps with a common fixed point,
- (ii) $|\Lambda^+(T)| = 1$ and each $g \in T$ is of the form $g(z) = az + b$ with $a, b \in \mathbb{C}$ and $|a| \geq 1$, or
- (iii) $|\Lambda^+(T)| = 2$ and each $g \in T$ is of the form $g(z) = az$ or $g(z) = a/z$ for some $a \in \mathbb{C} \setminus \{0\}$.

Definition We say that a semigroup S is *elementary* if there is a point $z \in \overline{\mathbb{B}^3}$ for which $S(z)$ is finite.

This definition is at odds with the definition used in [15], where a semigroup is described as elementary whenever $\Lambda^-(S)$ is finite. The definition used in [15] has two undesirable consequences. Firstly, it is possible for S to be elementary and S^{-1} not elementary; second, with this definition and when S happens to be a group, it is possible for S to be an elementary group but not an elementary semigroup. The former case happens if $S = \langle f, g \rangle$ where $f(z) = \frac{1}{3}z$ and $g(z) = \frac{1}{3}(z + 2)$. Then $\Lambda^+(S)$ is equal to the middle-thirds Cantor set and $\Lambda^-(S) = \{\infty\}$. The latter case, where S an elementary group but $\Lambda^-(S)$ is not finite, happens if, for example, S is the stabilizer in $\text{Aut}(\mathbb{B}^3)$ of some point $z \in \overline{\mathbb{C}}$. We shall call $\langle f, g \rangle$ the *Cantor semigroup*. The Cantor semigroup, which is given as an example in [15], is familiar and easy to analyse, and we shall use it again to illustrate new ideas.

The elementary semigroups are classified in the following theorem.

THEOREM 1.5. *Let S be an elementary semigroup. Then either*

- (i) *there is a point in \mathbb{B}^3 fixed by all elements of S ; or*
- (ii) *there is a point in $\overline{\mathbb{C}}$ fixed by all elements of S ; or*
- (iii) *there is a pair of points in $\overline{\mathbb{C}}$ fixed setwise by all elements of S .*

PROOF. Let S be an elementary semigroup with finite orbit X . We can assume that X is either contained in \mathbb{B}^3 or its ideal boundary. Any finite orbit is forward invariant, and so must be fixed setwise by each element in S .

One possibility is that $|X| \geq 3$ and X is contained in \mathbb{S}^2 . Since an element in $\text{Aut}(\mathbb{B}^3)$ is uniquely determined by its action on three distinct points in \mathbb{S}^2 , S itself must be finite, and in particular for any point $\zeta \in \mathbb{B}^3$ its orbit $S(\zeta)$ is finite. In this case we write $K = S(\zeta)$. Another possibility is that $X \subseteq \mathbb{B}^3$, in which case we set K equal to X . In both these cases we have a finite set $K \subseteq \mathbb{B}^3$ that is fixed setwise by each element in S . One possibility is that K is a single point. Otherwise we can define B to be the unique smallest closed hyperbolic ball containing K . We claim that S fixes the (hyperbolic) centre of B . For any $g \in S$ we have $K \subseteq B \cap g(B)$. If g does not fix B , then, by conjugating so that the origin $\mathbf{0}$ is the midpoint of the line from the centre of B to the centre of $g(B)$, it can be seen that $B \cap g(B)$ is contained in a hyperbolic ball of radius strictly less than the radius of B , contrary to our choice of B . It follows that no such g exists, as claimed. Hence if either $X \subseteq \mathbb{B}^3$, or both $|X| \geq 3$ and $X \subseteq \overline{\mathbb{C}}$, then we must have case (i). It remains to consider the case where $|X| \leq 2$ and $X \subseteq \overline{\mathbb{C}}$. We have already shown that each element of S fixes X setwise, hence we have case (ii) if $|X| = 1$ and case (iii) if $|X| = 2$. \square

The next observation follows immediately from the theorem.

COROLLARY 1.6. *The semigroup S is elementary if and only if S^{-1} is elementary.*

Below we give a finer classification of elementary semigroups, which follows easily from Theorems 1.4 and 1.5, and whose proof we omit.

PROPOSITION 1.7. *Any nontrivial elementary semigroup S is of one of the following types:*

- (a) $\Lambda^+ = \Lambda^- = \emptyset$ and S is a semigroup of hyperbolic rotations about a point in hyperbolic space.
- (b) $\Lambda^+ = \Lambda^- = \{x\}$ and S contains parabolic maps fixing x and possibly elliptic maps fixing x . These are the only possible maps in S .
- (c) $\Lambda^+ = \{\alpha\}$ and $\Lambda^- = \{\beta\}$ where $\alpha \neq \beta$. S contains one or more loxodromic maps with attracting fixed point α and repelling fixed point β . There may also be elliptic maps that fix both points.
- (d) $|\Lambda^+| = 1$ and $|\Lambda^-| > 2$. If $\Lambda^+ = \{\alpha\}$ then S contains loxodromic maps with attracting fixed point α and possibly parabolic or elliptic maps fixing α . The semigroup must contain at least two such loxodromic maps with different repelling fixed points.

- (e) $|\Lambda^-| = 1$ and $|\Lambda^+| > 2$. Here the inverse semigroup S^{-1} satisfies the preceding case.
- (f) $|\Lambda^+| = |\Lambda^-| = 2$. In this case $\Lambda^+ = \Lambda^- = \{x, y\}$ and S contains loxodromic maps and possibly elliptic maps fixing $\{x, y\}$ setwise. Both x and y must belong to both the set of attracting fixed points and the set of repelling fixed points of S .
- (g) For some $z \in \overline{\mathbb{C}}$ each element in S fixes z . Moreover $|\Lambda^+| > 2$ and $|\Lambda^-| > 2$.

In particular a semigroup that has at least one finite limit set is necessarily elementary. Case (g) represents the only elementary semigroups for which both limit sets are infinite. The distinction between elementary semigroups that have at least one finite limit set and those in case (g) is of significance. Accordingly we make the following definition.

Definition We say that an elementary semigroup S is of *finite type* if at least one of $\Lambda^+(S)$ and $\Lambda^-(S)$ is a finite set. An elementary semigroup that is not of finite type is of *infinite type*.

In Chapter 3 we shall consider subsemigroups of $\text{Aut}(\mathbb{D})$, and there we give a specialised version of the above proposition, so that we can more precisely understand such semigroups. Notice that if one limit set contains exactly 0 or 2 points, then so does the other limit set. Cases (d) and (e) are the only classes of elementary semigroup for which exactly one of the limit sets is perfect. If both limit sets are singletons, then they may or may not be equal. Case (a), where both limit sets are empty, is exactly case (i) in Theorem 1.5. Cases (c) and (f) correspond to case (iii) in Theorem 1.5. All other cases ((b),(d),(e), and (g)) correspond to case (ii) in Theorem 1.5.

An example of case (e) is given by the Cantor semigroup, that is the semigroup $\langle f, g \rangle$ where $f(z) = \frac{1}{3}z$ and $g(z) = \frac{1}{3}(z + 2)$.

In the theory of Kleinian groups, the limit set $\Lambda(G)$ of a Kleinian group G is the set of non-normality of G , that is, the complement of $\Lambda(G)$ is the largest open set upon which G is a normal family. Similarly if f is an analytic self-map of a Riemann surface, then the Julia set $J(f)$, defined to be the set of non-normality of the family $\{f^n \mid n \in \mathbb{N}\}$, is the smallest closed set that is fixed setwise by f . Moreover, when G is a nonelementary Kleinian group, the repelling fixed points of G are dense in $\Lambda(G)$, and, provided f is not

a parabolic Möbius transformation, the repelling fixed points of $\{f^n \mid n \in \mathbb{N}\}$ are dense in $J(f)$. For semigroups of Möbius transformations we have the following result.

THEOREM 1.8. *Suppose that S is a Möbius semigroup. Then:*

- (i) *provided S contains loxodromic elements, the set of repelling fixed points of S is dense in Λ^- ;*
- (ii) *the complement of Λ^- in \mathbb{S}^2 is the largest open set upon which S is a normal family; and*
- (iii) *if Λ^- contains at least three points then it is the smallest closed, backward invariant subset of the ideal boundary containing at least three points.*

PROOF. The first statement is given in [15, Theorem 2.4]. The second was essentially first proven in [1, Theorem 3.3], but also appears as [15, Theorem 2.6]. The final statement is [15, Remark 2.20]. The conclusion of statement (i) holds for all semigroups except those elementary semigroups that are not of case (b) in Proposition 1.7. \square

By applying Theorem 1.8 to the nonelementary semigroup S^{-1} , we observe that the attracting fixed points of S are dense in Λ^+ . Moreover, Λ^+ is the smallest closed, forward invariant subset of \mathbb{S}^2 containing at least three points. It follows that any forward invariant set containing at least three points is dense in Λ^+ , so that in particular Λ_c^+ and Λ_h^+ are dense in Λ^+ .

The concept of non-normality can be used to give a unified definition of the limit set $\Lambda(G)$ of a Kleinian group G and the Julia $J(f)$ set of an analytic function f . The above theorem tells us that we can also define the backward limit set Λ^- of a Möbius semigroup using normality. There are, however, important differences between the behaviour of Λ^- and of $\Lambda(G)$ and $J(f)$. The set $J(f)$ is fixed setwise by f , and $\Lambda(G)$ is fixed setwise by any element of G . In contrast for a semigroup S , in general neither limit set is fixed by all the elements in S .

We finish this chapter with some remarks on the dynamics of semigroups of analytic functions under composition. The study of semigroups of rational functions was introduced

in 1995 by Hinkkanen and Martin [20]. They define the Julia set of a semigroup of rational functions acting on $\overline{\mathbb{C}}$ to be the set of non-normality of the semigroup. In [20] the semigroup is taken to include at least one rational function that is not a Möbius transformation. The subject of semigroups of rational functions has since been further developed (see for example [19, 36–39]). Semigroups of transcendental entire functions acting on \mathbb{C} have also been studied; see, for example, [24].

CHAPTER 2

On Möbius semigroups

In this chapter we develop the theory of semigroups of Möbius transformations. In particular, we introduce the idea of a semidiscrete semigroup, which can be regarded as a generalisation of a discrete group. Some of our results on semidiscrete semigroups are analogues of known results on discrete groups, and it appears that the semidiscreteness property plays a role for semigroups somewhat similar to that of the discreteness property for groups.

1. The group and inverse free parts of a semigroup

The *group part* of a semigroup S is the set of elements in S whose inverses also lie in S . Hence the group part of S is the largest subset of S that is also a group, and is equal to $S \cap S^{-1}$. We call the complement of $S \cap S^{-1}$ in S the *inverse free part* of S , as it does not contain the identity. One possibility is that S itself is a group. In the other extreme, the group part of S might be empty, or equivalently $I \notin S$. This partition of a semigroup into its group part and inverse free part is particularly important because of the following lemma.

LEMMA 2.1. *Let S be a semigroup. If $g, h \in S$, and one of them belongs to $S \setminus S^{-1}$, then $gh \in S \setminus S^{-1}$.*

PROOF. We prove the contrapositive assertion that if $gh \in S \cap S^{-1}$ then $g, h \in S \cap S^{-1}$. If $gh \in S \cap S^{-1}$, then $f = (gh)^{-1}$ is an element of S . Therefore $g^{-1} = hf$ and $h^{-1} = fg$ both belong to S , as required. \square

It follows from Lemma 2.1 that the inverse free part of a semigroup is itself a semigroup. We define a *word* generated by a collection \mathcal{F} of Möbius transformations to be an expression of the form $f_1 \dots f_n$, where $n \in \mathbb{N}$ and each f_i belongs to \mathcal{F} . By regarding each

word as a composition of Möbius transformations, each word can be mapped to a Möbius transformation belonging to the semigroup S generated by \mathcal{F} . This mapping is certainly surjective. We say S is *freely generated by \mathcal{F}* (or simply *free*) if the mapping is also injective.

LEMMA 2.2. *Suppose that S is a semigroup generated by a set \mathcal{F} . If S has a nonempty group part, then that group is generated by a subset of \mathcal{F} .*

PROOF. Let \mathcal{G} denote the set of elements in \mathcal{F} that belong to the group part of S . By Lemma 2.1, a word in the generators \mathcal{F} represents an element of $S \cap S^{-1}$ if and only if all the letters in the word belong to \mathcal{G} . Therefore $S \cap S^{-1}$ is generated by \mathcal{G} . \square

In particular, the group part of a finitely-generated semigroup S is finitely-generated. In contrast, the inverse free part of S , that is, $S \setminus S^{-1}$, need not be finitely-generated, as the following example illustrates. Let f and g be two Möbius transformations that, as a group, generate a Schottky group. Consider the semigroup $S = \langle f, f^{-1}, g \rangle$. The set $S \setminus S^{-1}$ is exactly those Möbius transformations that arise from words generated by the set $\{f, f^{-1}, g\}$ that feature at least one occurrence of g . Since the group generated by f and g is freely generated, any finite set in $S \setminus S^{-1}$ cannot generate gf^n for every $n \in \mathbb{N}$. This means that even though S is finitely-generated, $S \setminus S^{-1}$ is not. In Section 2.4 of this chapter we shall consider how the limit sets of $S \setminus S^{-1}$ relate to the limit sets of S .

2. Semidiscrete semigroups

We now introduce an important property of semigroups.

Definition We say that a set X of Möbius transformations is *semidiscrete* if the identity element I is not an accumulation point of X in $\text{Aut}(\mathbb{B}^3)$.

As an example, consider the semigroup S generated by $f(z) = 2z$ and $g(z) = \frac{1}{2}z + 1$. These are loxodromic Möbius transformations such that the attracting fixed point of f equals the repelling fixed point of g , that is $\alpha_f = \beta_g$. This semigroup is not discrete because

$$g^n f^n(z) = z + 2 - \frac{1}{2^{n-1}} \rightarrow z + 2 \quad \text{as } n \rightarrow \infty.$$

However, it is semidiscrete because each element of S has the form $z \mapsto 2^n z + b$, where $n \in \mathbb{Z}$ and $b \in \mathbb{R}$, and it is easy to check that if $n = 0$ then $b > 1$.

In the theory of groups there are several properties that are equivalent to discreteness. Accordingly, we now consider properties of semigroups that are equivalent to being semidiscrete. The action of a semigroup S on \mathbb{B}^3 is said to be *properly discontinuous* if for each point ζ in \mathbb{B}^3 there is a neighbourhood U of ζ such that $g(U) \cap U \neq \emptyset$ for only finitely many elements g of S . When S is a group, it is discrete if and only if its action on \mathbb{B}^3 is properly discontinuous, and this is so precisely when the S -orbit of any point in \mathbb{B}^3 is locally finite. The next theorem is a comparable result for semigroups.

THEOREM 2.3. *Let S be a semigroup. The following statements are equivalent:*

- (i) *S is semidiscrete;*
- (ii) *the action of S on \mathbb{B}^3 is properly discontinuous;*
- (iii) *the S -orbit of any point ζ in \mathbb{B}^3 does not accumulate at ζ .*

We omit the elementary proof, which is similar to proofs of analogous theorems from the theory of Kleinian groups.

The action of a semigroup S on \mathbb{B}^3 is said to be *strongly discontinuous* if for each point ζ in \mathbb{B}^3 there is a neighbourhood U of ζ such that $g(U) \cap U = \emptyset$ for every element g of S . This definition is close to the definition of a discontinuous action in the theory of Fuchsian groups, but there the intersection $g(U) \cap U$ is only required to be empty for elements of the group other than the identity.

THEOREM 2.4. *Let S be a semigroup. The following statements are equivalent:*

- (i) *S is semidiscrete and inverse free;*
- (ii) *the action of S on \mathbb{B}^3 is strongly discontinuous;*
- (iii) *the S -orbit of any point ζ in \mathbb{B}^3 stays a positive distance away from ζ .*

Again, the proof is straightforward, and omitted.

3. Schottky semigroups

Here we introduce a large class of finitely-generated, inverse free semidiscrete semigroups. We first give a condition which ensures that a semigroup is inverse free and semidiscrete.

THEOREM 2.5. *Let K be a closed subset of \mathbb{S}^2 and let \mathcal{F} be a finite subset of $\text{Aut}(\mathbb{B}^3)$ comprised of transformations that map K within itself. Further suppose that the elements of \mathcal{F} that fix K setwise generate a finite group. Then the semigroup S generated by \mathcal{F} is semidiscrete. If no element in \mathcal{F} fixes K setwise, then S is also inverse free.*

PROOF. First we prove that the semigroup S generated by \mathcal{F} is semidiscrete. Suppose, on the contrary, that S is not semidiscrete; then there is a sequence of distinct elements g_n of S that converges to the identity map I . Let \mathcal{F}_0 denote those elements in \mathcal{F} that fix K setwise, and let \mathcal{F}_1 denote those elements in \mathcal{F} that map K strictly within itself. Let G be the group generated by \mathcal{F}_0 . Each map g_n can be written as a composition of elements of \mathcal{F} . One possibility is that infinitely many g_n lie in G . Then since G is finite we have $g_n = I$ for infinitely many n , contrary to our assumption that the elements in the sequence g_n are distinct. Otherwise we may pass to a subsequence such that we can write $g_n = pfq_n$, for $n = 1, 2, \dots$, where $f \in \mathcal{F}_1$, $q_n \in S$ and $p \in G \cup \{I\}$. Then $g_n(K) = p(f(q_n(K))) \subseteq p(f(K))$. Since $f(K)$ is a proper subset of K , so is $pf(K)$. It follows that g_n cannot accumulate at the identity. Hence S is semidiscrete.

To show the contrapositive of the final statement, suppose that S contains the identity. Then $I = f_1 \cdots f_m$ for some $f_1, \dots, f_m \in \mathcal{F}$ and positive integer m . Suppose towards contradiction that some f_i lies in \mathcal{F}_1 . Then $f_1 \cdots f_m(K) \subseteq f_1 \cdots f_i(K) \subsetneq f_1 \cdots f_{i-1}(K)$, so that $f_1 \cdots f_m$ cannot be the identity map. Hence each f_i lies in \mathcal{F}_0 , and in particular \mathcal{F}_0 is nonempty. \square

Suppose now that K is the union of a finite collection of disjoint, closed discs in \mathbb{S}^2 , and let \mathcal{F} be a finite subset of $\text{Aut}(\mathbb{B}^3)$ made up of transformations that map K strictly within itself. A *Schottky semigroup* is a semigroup generated by such a set \mathcal{F} . We use this terminology because Schottky groups contain many Schottky semigroups; for example, if f and g generate as a group a Schottky group, then f and g generate as a semigroup a Schottky semigroup.

We now give an example of a finitely-generated inverse free semidiscrete semigroup that is not a Schottky semigroup. In this example, as usual, we denote the attracting and

repelling fixed points of a loxodromic element g of $\text{Aut}(\mathbb{B}^3)$ by α_g and β_g , respectively. Let f, g and h be loxodromic maps in $\text{Aut}(\mathbb{D})$ that generate *as a group* a Schottky group (not a Schottky semigroup). Choose these maps in such a way that α_f and α_h lie in different components of $\mathbb{S}^1 \setminus \{\alpha_g, \beta_g\}$. Let $q = fg^{-1}f^{-1}$, and define $S = \langle f, g, h, q \rangle$ (see Figure 2.1). This semigroup lies in a discrete group, so it is semidiscrete. To see that S is inverse free, suppose, in order to reach a contradiction, that $w_1 \cdots w_n = I$, where w_i is a positive power of either f, g, h or q for $i = 1, \dots, n$. By thinking of $w_1 \cdots w_n$ as a word in f, g and h , we see that w_i cannot equal f or h (because the sum of the powers of each of f, g and h in this word must be 0, as those three maps generate a free group). Therefore each map w_i is equal to either g^m or $fg^{-m}f^{-1}$, for some positive integer m , so clearly it is not possible for $w_1 \cdots w_n$ to equal the identity after all.

It remains to prove that there is not a finite collection of disjoint, closed intervals in \mathbb{S}^1 whose union K is mapped strictly within itself by each element of S . Suppose there is such a set K . Then it must contain $\Lambda^+(S)$, so, in particular, it contains α_g . Furthermore, it contains $g^n(\alpha_f)$ and $g^n(\alpha_h)$ for each positive integer n . These points accumulate on either side of α_g (our initial choice of f, g and h ensures that this is so). It follows that α_g is an interior point of K . Now, $q = fg^{-1}f^{-1}$, so $\beta_q = f(\alpha_g)$, which implies that β_q is also an interior point of K . However, this is impossible because $q(K) \subseteq K$. Therefore S is not a Schottky semigroup.

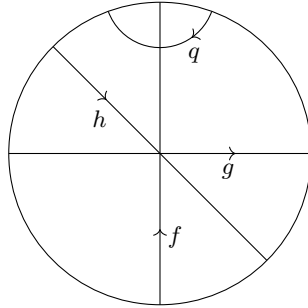


FIGURE 2.1. The axes of f, g, q, h

An important feature of this example is that the generating set is not unique, because $S = \langle fg, g, h, q \rangle$, as one can easily verify. Although S is contained in a Schottky group,

this property is not mandatory for a finitely-generated, semidiscrete and inverse free semigroup whose generating set is not unique. Indeed, by appending an appropriate Möbius transformation onto S , we can generate a larger semigroup with these same properties that is not contained in a Schottky group. More specifically, we can use the construction outlined in Chapter 2, Section 6 to choose an element $t \in \text{Aut}(\mathbb{D})$ such that $T = \langle f, g, h, q, t \rangle$ is also semidiscrete, inverse free and finitely-generated, but not uniquely generated by $\{f, g, h, q, t\}$, nor contained within a discrete group. The problem of determining whether or not a particular generating set for a semigroup is unique will be dealt with in Theorem 4.25 and the surrounding discussion.

4. Composition sequences

Recall that a sequence G_n of Möbius transformations is an escaping sequence if, for some point ζ in \mathbb{B}^3 , the orbit $G_n(\zeta)$ does not accumulate at any point in \mathbb{B}^3 . There are various equivalent ways of describing escaping sequences, one of which is captured in the following standard lemma (we omit the elementary proof).

LEMMA 2.6. *A sequence of Möbius transformations is an escaping sequence if and only if it does not contain a subsequence that converges to a Möbius transformation.*

The statement that a sequence G_n of Möbius transformations converges uniformly to another Möbius transformation G is equivalent to $\sigma(G_n, G) \rightarrow 0$ as $n \rightarrow \infty$, where σ is the metric of uniform convergence introduced in Chapter 1. The relationship between composition sequences and semigroups is a central theme of this thesis. The key to this relationship is the following simple theorem, which characterises when every composition sequence generated by S converges.

THEOREM 2.7. *Let $S = \langle \mathcal{F} \rangle$ be a semigroup. The following are equivalent:*

- (i) *S is semidiscrete and inverse free.*
- (ii) *Every composition sequence generated by \mathcal{F} is an escaping sequence.*
- (iii) *Every composition sequence generated by S is an escaping sequence.*

PROOF. We first show that (iii) implies (i). Suppose that S is not both semidiscrete and inverse free. Then either the identity transformation I is an accumulation point of

S in $\text{Aut}(\mathbb{B}^3)$, or else it belongs to S . In other words, $I \in \overline{S}$. Hence there is a sequence g_n in S that converges to I . By restricting to a subsequence of g_n , we can assume that $\sum \sigma(g_n, I) < +\infty$. Now define $G_n = g_1 \cdots g_n$ for $n \in \mathbb{N}$. Using the right-invariance of σ we see that

$$\sigma(G_n^{-1}, G_{n-1}^{-1}) = \sigma(G_n^{-1}G_n, G_{n-1}^{-1}G_n) = \sigma(I, g_n).$$

Therefore $\sum \sigma(G_n^{-1}, G_{n-1}^{-1}) < +\infty$, which implies that G_n^{-1} is a Cauchy sequence. Hence G_n is a Cauchy sequence too. As σ is a complete metric on $\text{Aut}(\mathbb{B}^3)$, the sequence G_n converges. Hence by Lemma 2.6, G_n is not an escaping sequence, and so we have found a composition sequence generated by S that is not an escaping sequence.

To see (i) implies (iii), suppose that there are maps g_n in S such that the composition sequence $G_n = g_1 \cdots g_n$ generated by S is not an escaping sequence. Then, by Lemma 2.6, there is a subsequence G_{n_k} that converges to a Möbius transformation G . It follows that $G_{n_{k-1}}^{-1}G_{n_k} \rightarrow I$ as $k \rightarrow \infty$. But

$$G_{n_{k-1}}^{-1}G_{n_k} = (g_1 \cdots g_{n_{k-1}})^{-1}(g_1 \cdots g_{n_k}) = g_{n_{k-1}+1} \cdots g_{n_k},$$

so this sequence lies in S . It follows that $I \in \overline{S}$, so S is not both semidiscrete and inverse free.

Since each composition sequence generated by S is a subsequence of some composition sequence generated by \mathcal{F} , we see that (ii) implies (iii). On the other hand $\mathcal{F} \subseteq S$, and so (iii) implies (ii). \square

We say that a sequence of Möbius transformations is *discrete* if the set of transformations that make up the sequence is a discrete subset of $\text{Aut}(\mathbb{B}^3)$. Although escaping sequences are all discrete, by definition, the converse does not hold; for example, the trivial sequence I, I, I, \dots is discrete, but it is not an escaping sequence.

The preceding theorem gives a condition in terms of composition sequences for a semigroup to be both semidiscrete and inverse free. The next, similar theorem gives a condition in terms of composition sequences for a semigroup to be semidiscrete. The proof is similar, so we only sketch the details.

THEOREM 2.8. *Let $S = \langle \mathcal{F} \rangle$ be a semigroup. The following are equivalent:*

- (i) S is semidiscrete.

- (ii) *Every composition sequence generated by \mathcal{F} is discrete.*
- (iii) *Every composition sequence generated by S is discrete.*

PROOF. We first show that (iii) implies (i). Suppose that S is not semidiscrete. Then there is a sequence of *distinct* transformations g_n from S that converges to I . As before, we can define a composition sequence $G_n = g_1 \cdots g_n$ for $n = 1, 2, \dots$, and, providing that g_n converges to I sufficiently quickly, we see that G_n converges to some Möbius transformation G . Because the maps g_n are distinct, it follows that $G_n \neq G$ for infinitely many n . Hence G_n is not discrete.

To see that (i) implies (iii), suppose that there are maps g_n in S such that the composition sequence $G_n = g_1 \cdots g_n$ generated by S is not discrete. Then we can choose a subsequence G_{n_k} consisting of distinct maps that converges to a Möbius transformation G . Therefore $G_{n_{k-1}}^{-1} G_{n_k} \rightarrow I$ as $k \rightarrow \infty$, where $G_{n_{k-1}}^{-1} G_{n_k} \in S \setminus \{I\}$. This implies that S is not semidiscrete.

Since each composition sequence generated by S is a subsequence of some composition sequence generated by \mathcal{F} , we see that (ii) implies (iii). Conversely since $\mathcal{F} \subseteq S$ it follows that (iii) implies (ii). \square

5. Covering regions

In this section we introduce the concept of a covering region for a semigroup. These appear to play a role for semidiscrete semigroups somewhat similar to the role played by fundamental regions for discrete groups. A *covering region* for a semigroup S is a closed subset D of \mathbb{B}^3 with nonempty interior such that

$$\bigcup_{g \in S \cup \{I\}} g(D) = \mathbb{B}^3.$$

Trivially, \mathbb{B}^3 itself is a covering region for S . Let us denote the interior of a set X by $\text{int}(X)$. We say that a covering region D is a *fundamental region* for S if it satisfies the additional property $\text{int}(D) \cap \text{int}(g(D)) = \emptyset$ whenever g is a nonidentity element of S . This definition coincides with the usual definition of a fundamental region when S is a Kleinian group.

The next theorem due to Bell [10, Theorem 3] says that if a semigroup has a fundamental region, then it is in fact a Kleinian group. Although we do not use the next theorem again, we include it because [10] is difficult to obtain, and the proof presented there is excessively complicated.

THEOREM 2.9. *Suppose that D is a fundamental region for a semigroup S such that $\overline{\text{int}(D)} = D$. Then S is a Kleinian group with fundamental region D .*

PROOF. Let $p \in \text{int}(D)$ and choose any element g of S . As D is a fundamental region, there is another element h of $S \cup \{I\}$ such that $g^{-1}(p) \in h(D)$. Hence $p \in gh(D)$, and since $p \in \text{int}(D)$ and there are points in $gh(\text{int}(D))$ that accumulate at p (using that $\overline{\text{int}(D)} = D$) we see that $\text{int}(D) \cap gh(\text{int}(D)) \neq \emptyset$. Therefore $gh = I$, and so $h = g^{-1}$. Since g was arbitrarily chosen, S is a group with fundamental region D . Hence it must be a Kleinian group because it has a fundamental region. \square

We remark that in the theorem above, instead of assuming that $\overline{\text{int}(D)} = D$, we can suppose that S is countable. To see this we argue as follows: First note that D must have non-empty interior, for otherwise D is nowhere dense (as D is closed), in which case $S(D)$ cannot cover \mathbb{B}^3 by the Baire category theorem. If g is an element of S then $g^{-1}(\text{int}(D))$ meets $S(D) = S(\partial D) \cup S(\text{int}(D))$. Note that ∂D is nowhere dense since D is closed. Since S is countable the Baire category theorem tells us that $S(\partial D)$ has empty interior. Hence $S(\partial D)$ does not cover $g^{-1}(\text{int}(D))$, and so $g^{-1}(\text{int}(D)) \cap h(\text{int}(D)) \neq \emptyset$ for some $h \in S$. Hence $\text{int}(D) \cap gh(\text{int}(D)) \neq \emptyset$ and it follows that $gh = I$, that is $h = g^{-1}$.

We now consider covering regions that are not necessarily fundamental regions. In the following lemma we refer to the inverse free part of a semigroup S , which is the set $S \setminus S^{-1}$.

LEMMA 2.10. *If D is a covering region for a semigroup S that is not a group, then D is also a covering region for the inverse free part of S .*

PROOF. Choose any point p in \mathbb{B}^3 and let $f \in S \setminus S^{-1}$. Then there is an element g of $S \cup \{I\}$ such that $f^{-1}(p) \in g(D)$. Hence $p \in fg(D)$. By Lemma 2.1, $fg \in S \setminus S^{-1}$. As p was chosen arbitrarily it follows that D is a covering region for $S \setminus S^{-1}$. \square

We now give the main result of this section, which says that any semidiscrete semigroup with a bounded covering region is in fact a group. The proof of this theorem is a good example of how composition sequences themselves can be used as a tool for studying semigroups.

THEOREM 2.11. *Any semidiscrete semigroup that has a bounded covering region is a Kleinian group.*

PROOF. Let S be a semidiscrete semigroup and choose any point ζ in \mathbb{B}^3 . As S has a bounded covering region, we can choose a suitably large open hyperbolic ball D , with (hyperbolic) radius r and centre ζ , such that

$$\bigcup_{g \in S \cup \{I\}} g(D) = \mathbb{B}^3.$$

As ∂D is compact, there is a finite subset T of S such that

$$\partial D \subseteq \bigcup_{g \in T} g(D).$$

We suppose T has been chosen with no redundancy, so that $\rho(g(\zeta), \zeta) \leq 2r$ for each $g \in T$.

We construct a composition sequence $G_n = g_1 \cdots g_n$ generated by T as follows.

- (i) If $n > 1$ and $\zeta \in G_{n-1}(\overline{D})$, then choose g_n arbitrarily from T .
- (ii) Otherwise, choose g_n from T such that $\rho(G_n(\overline{D}), \zeta)$ is minimised.

Let $\rho_n = \rho(G_n(\overline{D}), \zeta)$ and suppose $n \geq 2$. If g_n is chosen according to (i), then $\rho(G_{n-1}(\zeta), \zeta) \leq r$, so

$$\rho(G_n(\zeta), \zeta) \leq \rho(G_n(\zeta), G_{n-1}(\zeta)) + \rho(G_{n-1}(\zeta), \zeta) = \rho(g_n(\zeta), \zeta) + \rho(G_{n-1}(\zeta), \zeta) \leq 3r.$$

Hence $\rho_n \leq 3r$. Suppose now that g_n is chosen according to (ii). Note that the collection $\{G_{n-1}g(D) \mid g \in T\}$ covers $G_{n-1}(\partial D)$. Therefore

$$\rho(G_{n-1}(\overline{D}), \zeta) = \rho(G_{n-1}(\partial D), \zeta) \geq \rho(G_n(\overline{D}), \zeta);$$

that is, $\rho_{n-1} \geq \rho_n$.

It follows that the sequence ρ_n is bounded, and so G_n is not an escaping sequence. Theorem 2.7 now tells us that S is not inverse free. We know from Lemma 2.10 that the inverse free part of S , if nonempty, also has D as a covering region. But applying the

above argument with $S \setminus S^{-1}$ replacing S gives a contradiction, because $S \setminus S^{-1}$ is inverse free. Therefore $S \setminus S^{-1}$ is empty, and so S is a semidiscrete group, that is, a Kleinian group. \square

The converse is false: if S is a Kleinian group it may have no bounded covering region, regardless of whether or not its limit set is the full ideal boundary.

Given a semidiscrete semigroup S and a point w in \mathbb{B}^3 that is not fixed by any element of $S \setminus \{I\}$, we define the *Dirichlet region* for S centred at w to be the set

$$D_w(S) = \{z \in \mathbb{B}^3 \mid \rho(z, w) \leq \rho(z, g(w)) \text{ for all } g \text{ in } S \setminus \{I\}\}.$$

This is the same definition of a Dirichlet region used in the theory of Kleinian groups. Because the points $S(w)$ do not accumulate at w or contain w , the Dirichlet region centred at w always contains some open neighbourhood of w . We next verify that Dirichlet regions of semidiscrete semigroups are indeed covering regions.

THEOREM 2.12. *Let S be a semidiscrete semigroup and let w be a point in \mathbb{B}^3 that is not fixed by any element of $S \setminus \{I\}$. Then $D_w(S)$ is a covering region for S .*

PROOF. The set $D_w(S)$ is closed because it is an intersection of closed sets (similarly it is also convex, although we do not use this fact). It has nonempty interior because, by Theorem 2.3, the orbit of w under S does not accumulate at w . To verify the covering property, suppose towards contradiction that there is a point z in \mathbb{B}^3 that does not lie in $h(D_w(S))$ for any element h of $S \cup \{I\}$. Then $h^{-1}(z) \notin D_w(S)$, which implies that there is a map g in $S \cup \{I\}$ that satisfies $\rho(h^{-1}(z), w) > \rho(h^{-1}(z), g(w))$. In other words,

$$(3) \quad \text{for all } h \in S \cup \{I\} \text{ there exists } g \in S \cup \{I\} \text{ such that } \rho(z, h(w)) > \rho(z, hg(w)).$$

We use (3) to recursively define a composition sequence $F_n = f_1 \cdots f_n$, where $f_i \in S \cup \{I\}$. Let $f_1 = I$. If f_1, \dots, f_{n-1} have been defined, then, using (3), we let f_n be an element of $S \cup \{I\}$ that satisfies $\rho(z, F_{n-1}(w)) > \rho(z, F_{n-1}f_n(w))$. The resulting composition sequence F_n satisfies

$$\rho(z, w) > \rho(z, F_1(w)) > \rho(z, F_2(w)) > \cdots.$$

As S is semidiscrete, Theorem 2.8 tells us that the composition sequence F_n is discrete. But the sequence $F_n(w)$ is bounded, so it must have a constant subsequence. This contradicts the sequence of strict inequalities above. Hence, contrary to our earlier assumption,

$$z \in \bigcup_{h \in S \cup \{I\}} h(D_w(S)),$$

and so $D_w(S)$ is a covering region for S . \square

We remark that for a semidiscrete semigroup it is always possible to choose a point $w \in \mathbb{B}^3$ that is not fixed by any element of $S \setminus \{I\}$. The set of points in \mathbb{B}^3 that are fixed by some element of $S \setminus \{I\}$ are necessarily fixed points of elliptic transformations that lie in S . Since S is semidiscrete and inverse free the elliptic transformations in S form a discrete group, and in particular there are countably many of them. Hence the set of $w \in \mathbb{B}^3$ fixed by an element of $S \setminus \{I\}$ is contained in a countable union of geodesics of \mathbb{B}^3 . It follows that one can always define a Dirichlet region for semidiscrete semigroups.

Recall from Chapter 1 that the forward horospherical limit set of S , $\Lambda_h^+(S)$, is the set of points x in $\Lambda^+(S)$ such that, for any point ζ in \mathbb{B}^3 , the orbit $S(\zeta)$ meets every horoball based at x . The remainder of this section is concerned with proving that if the forward horospherical limit set of a semidiscrete semigroup is the entire ideal boundary, then every Dirichlet region of S is bounded. As we do this, it is convenient to work with the upper half-space model of hyperbolic space, chiefly because in this model, a horoball based at ∞ takes a convenient form. Recall that a point $z \in \mathbb{H}^3$ can be written in the form $z = x + t\mathbf{j}$ where $x \in \mathbb{C}$ and $t > 0$. We define the *height* of z to be $\text{ht}[z] = t$. Given x in \mathbb{C} and w in \mathbb{H}^3 , let $H_x(w)$ denote the open horoball that is based at x and tangent to w . So, working in the upper half-space model, $H_\infty(w) = \{z \in \mathbb{H}^3 \mid \text{ht}[z] > \text{ht}[w]\}$. Let $\gamma_z = \{z + t\mathbf{j} \mid t \geq 0\}$ denote the vertical geodesic from z to ∞ . Then $\overline{\gamma_z} = \gamma_z \cup \{\infty\}$. Given distinct points u and v in \mathbb{H}^3 , we define

$$K(u, v) = \{z \in \mathbb{H}^3 \mid \rho(z, u) \leq \rho(z, v)\}.$$

Note that $\overline{K(u, v)}$ is the closure of $K(u, v)$ in $\overline{\mathbb{R}^3}$ (it is not the same as $K(u, v)$).

The next lemma is an elementary exercise in hyperbolic geometry.

LEMMA 2.13. *Suppose that u and v are distinct points in \mathbb{H}^3 and $\text{ht}[u] < \text{ht}[v]$. Then*

$$\overline{K(u, v)} \cap \overline{\gamma_v} = \emptyset.$$

We use the lemma to prove the following theorem, which we shall need in Chapter 3, and helps us relate Dirichlet regions to horospherical limit sets.

THEOREM 2.14. *Let S be a semidiscrete semigroup, $x \in \overline{\mathbb{C}}$ and $w \in \mathbb{H}^3$. Then $x \in \overline{D_w(S)}$ if and only if the S -orbit of w does not meet $H_x(w)$.*

PROOF. By conjugating, we can assume that $x = \infty$ and $w = \mathbf{j}$. First suppose that the S -orbit of \mathbf{j} meets $H_\infty(\mathbf{j})$. Then there is an element g of S such that $g(\mathbf{j}) \in H_\infty(\mathbf{j})$, and so $\text{ht}[g(\mathbf{j})] > 1$. By Lemma 2.13, $\infty \notin \overline{K(\mathbf{j}, g(\mathbf{j}))}$. Since $\overline{D_{\mathbf{j}}(S)}$ is contained in $\overline{K(\mathbf{j}, g(\mathbf{j}))}$, we see that $\infty \notin \overline{D_{\mathbf{j}}(S)}$.

Conversely, suppose that $g(\mathbf{j}) \notin H_\infty(\mathbf{j})$ for every element g of S . Then for each map g , we have $\gamma_{\mathbf{j}} \subseteq K(\mathbf{j}, g(\mathbf{j}))$. Therefore $\gamma_{\mathbf{j}} \subseteq D_{\mathbf{j}}(S)$, so $\infty \in \overline{D_{\mathbf{j}}(S)}$. \square

An immediate consequence of the theorem is the following important corollary.

COROLLARY 2.15. *Let S be a semidiscrete semigroup. If $\Lambda_h^+(S) = \overline{\mathbb{C}}$, then S is a Kleinian group and every Dirichlet region of S is bounded in \mathbb{H}^3 .*

PROOF. Let $D_w(S)$ be a Dirichlet region for S . We are given that $\Lambda_h^+(S) = \overline{\mathbb{C}}$, so the orbit $S(w)$ meets every horoball in \mathbb{H}^3 . Theorem 2.14 now tells us that $\overline{D_w(S)} \cap \overline{\mathbb{C}} = \emptyset$. Therefore $D_w(S)$ is bounded in \mathbb{H}^3 , and we infer from Theorem 2.11 that S is a Kleinian group. \square

In the theory of Kleinian groups there are various theorems that relate the group to the geometry of its Dirichlet region. For example, any Dirichlet region of a Fuchsian group has finitely many sides (that is, maximal subsets of the region's boundary that are contained within geodesic segments) precisely when the group is finitely-generated. It is natural to ask whether results of this type exist for semidiscrete semigroups. Corollary 2.15 is one such result and in Theorem 2.17 we use the Dirichlet region to characterise exactly when the forward limit set is equal to the full ideal boundary, at least for countable, semidiscrete and inverse free semigroups.

For the rest of this chapter it is convenient to work in the ball model. For a semidiscrete semigroup S , and any $w \in \mathbb{B}^3$ that is not fixed by any element of S , we define the *ideal boundary of $D_w(S)$* , which we denote by $e_w(S)$, to be $\overline{D_w(S)} \cap \mathbb{S}^2$, where the closure is taken with respect to the Euclidean metric. As usual we abbreviate $e_w(S)$ to e_w whenever the semigroup is unambiguous. We note that e_w is a closed subset of \mathbb{S}^2 . We shall show that, under certain assumptions, e_w has a non-empty interior precisely when the forward limit set is not the full ideal boundary. Towards this, we need an initial lemma, the proof of which makes use of composition sequences in an essential way. We shall need to consider the set of points in Λ^+ that are the limit points of some composition sequence generated by S . We denote the set of such points by Λ_q^+ , and note that Λ_q^+ is forward invariant, and so by Theorem 1.8 Λ_q^+ is a dense subset of Λ^+ whenever S is nonelementary.

LEMMA 2.16. *Suppose that S is a countable, semidiscrete and inverse free semigroup. Then for all $w \in \mathbb{B}^3$ we have*

$$\mathbb{S}^2 \setminus \Lambda_q^+ \subseteq S(e_w).$$

PROOF. First note that since S is semidiscrete and inverse free, no point in \mathbb{B}^3 is fixed by any element of S , and so e_w , the ideal boundary of $D_w(S)$, is defined. For a point $x \in \mathbb{S}^2$ recall that $H_x(w)$ is the horoball based at x such that $w \in \mathbb{B}^3$ lies on its boundary. It suffices to show that for all $x \in \mathbb{S}^2 \setminus \Lambda_q^+$ and all $w \in \mathbb{B}^3$, there exists $g \in S$ such that $g^{-1}(x) \in e_w$. Suppose towards contradiction that this is false, that is, there exists some $w \in \mathbb{B}^3$ and $x \in \mathbb{S}^2 \setminus \Lambda_q^+$ such that $g^{-1}(x) \notin e_w$ for all $g \in S$. Theorem 2.14 tells us that $x \in e_w$ if and only if $S(w) \cap H_x(w) = \emptyset$. Hence

$$S(w) \cap H_{g^{-1}(x)}(w) \neq \emptyset,$$

or equivalently

$$(4) \quad gS(w) \cap H_x(g(w)) \neq \emptyset,$$

for all $g \in S$. Now choose any $f_1 \in S$. Using (4) and setting $g = f_1$, there exists $f_2 \in S$ such that

$$f_1 f_2(w) \in H_x(f_1(w)).$$

Again appealing to (4) with $g = f_1 f_2$, there exists $f_3 \in S$ such that

$$f_1 f_2 f_3(w) \in H_x(f_1 f_2(w)).$$

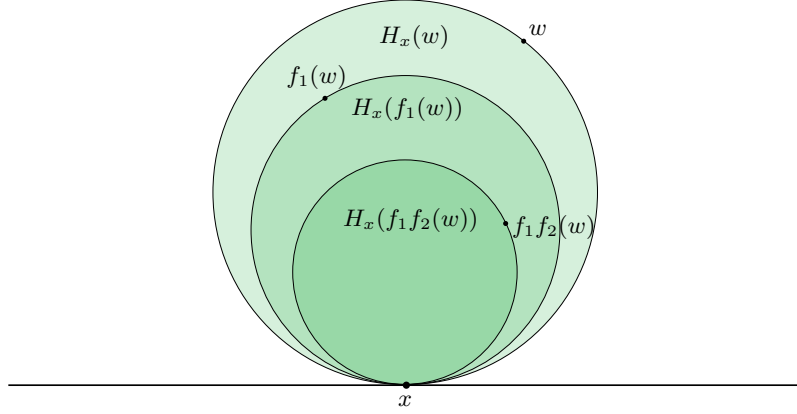


FIGURE 2.2. The sequence $F_n(w)$

Continuing in this way (see Figure 2.2) we obtain a sequence f_n in S such that

$$F_{n+1}(w) \in H_x(F_n(w)),$$

for each $n \in \mathbb{N}$, where $F_n = f_1 \cdots f_n$.

Since S is semidiscrete and inverse free, Theorem 2.7 tells us that the sequence $F_n(w)$ accumulates only on the ideal boundary. Since each point in $F_n(w)$ belongs to the horoball $H_x(f_1(w))$, the sequence of points $F_n(w)$ converges to x . But this contradicts the assumption that $x \notin \Lambda_q^+$; hence, for any $w \in \mathbb{B}^3$, we have $x \in g(e_w)$ for some $g \in S$. Since x was chosen arbitrarily in the complement of Λ_q^+ we have shown

$$\mathbb{S}^2 \setminus \Lambda_q^+ \subseteq S(e_w),$$

as required. □

Theorem 2.11 says that if a semidiscrete semigroup has a bounded covering domain, then it is a group. The next theorem is of a similar flavour.

THEOREM 2.17. *Suppose S is a countable, semidiscrete and inverse free semigroup. Then for any $w \in \mathbb{B}^3$ the set e_w has empty interior as a subspace of \mathbb{S}^2 if and only if $\Lambda^+ = \mathbb{S}^2$.*

PROOF. Suppose e_w contains an open set, U say. Since Λ_h^+ is both dense in Λ^+ and disjoint from e_w it follows that Λ^+ does not meet U , and so $\Lambda^+ \neq \mathbb{S}^2$.

For the converse, suppose that $\Lambda^+ \neq \mathbb{S}^2$. Note that a closed set is nowhere dense if and only if it does not contain an open set. Towards contradiction suppose that e_w is nowhere dense. Since S is countable $S(e_w)$ is a meagre set, that is, a countable union of nowhere dense sets. It follows from the Baire category theorem that $S(e_w)$ has empty interior. Now by Lemma 2.16 we have

$$\text{int}(\mathbb{S}^2 \setminus \Lambda_q^+) \subseteq \text{int}(S(e_w)) = \emptyset,$$

and so

$$\mathbb{S}^2 \setminus \overline{\Lambda_q^+} = \emptyset,$$

where the interiors and closures are taken with respect to the Euclidean topology on \mathbb{S}^2 . But since Λ_q^+ is dense in Λ^+ it follows that $\Lambda^+ = \mathbb{S}^2$, contrary to our assumption. Hence e_w cannot be nowhere dense after all. \square

It follows that for any countable, semidiscrete and inverse free semigroup S , if w and w' are points in \mathbb{B}^3 that are not fixed by any element of S , then e_w has empty interior if and only if $e_{w'}$ also has empty interior. In fact we have the following theorem, due to Edward Crane.

THEOREM 2.18. *Suppose S is a countable, semidiscrete and inverse free semigroup, and w and w' are points in \mathbb{B}^3 . Then e_w is empty if and only if $e_{w'}$ is also empty.*

PROOF. First suppose that e_w is empty, and so D_w is a bounded covering region of S . Since D_w is a compact fundamental region, it follows from Theorem 2.11 that S is a Kleinian group. Let r be the diameter of the quotient orbifold $M = \mathbb{B}^3/S$, which is finite because D_w is compact. Let $\pi : \mathbb{B}^3 \rightarrow M$ be the quotient map. Then for any point $w' \in \mathbb{B}^3$ and any $x \in \mathbb{B}^3$ such that $\rho(x, w') > r$, we can lift a shortest path in M from $\pi(x)$ to $\pi(w')$ to a path in \mathbb{B}^3 of the same length (at most r), starting at x . The lifted path finishes at some $g(w')$ for $g \in S \setminus \{I\}$, and $\rho(x, w') > r \geq \rho(x, g(w'))$, so $x \notin D_{w'}(S)$. Hence $D_{w'}(S)$ is bounded, and $e_{w'} = \emptyset$ as required. \square

6. Constructing new semidiscrete semigroups from old

We now describe a technique based on one of the well known combination theorems of Klein, given in the introduction of [30] and which we now state.

THEOREM 2.19. *Suppose G_1 and G_2 are two finitely-generated Kleinian groups with fundamental regions D_1 and D_2 respectively. Further suppose that the interior of D_1 contains the boundary and exterior of D_2 , and the interior of D_2 contains the boundary and exterior of D_1 . Then the group G generated by $G_1 \cup G_2$ is discrete and $D_1 \cap D_2$ is a fundamental region for G .*

Given distinct points u and v in \mathbb{B}^3 , recall that $K(u, v) = \{z \in \mathbb{B}^3 : \rho(z, u) \leq \rho(z, v)\}$. Notice that $g(K(u, v)) = K(g(u), g(v))$ for any Möbius transformation g . If S is a semidiscrete semigroup, and w is a point of \mathbb{B}^3 that is not fixed by any element of S , then we can express the Dirichlet region for S centred at w as

$$D_w(S) = \bigcap_{g \in S \setminus \{I\}} K(w, g(w)).$$

Then

$$\mathbb{B}^3 \setminus D_w(S) = \bigcup_{g \in S \setminus \{I\}} K(g(w), w)^\circ,$$

where $K(g(w), w)^\circ = \{z \in \mathbb{B}^3 \mid \rho(z, g(w)) < \rho(z, w)\}$ is the interior of $K(g(w), w)$.

THEOREM 2.20. *Suppose that S_1 and S_2 are semidiscrete semigroups and w is a point in \mathbb{B}^3 that is not fixed by any nontrivial element of S_1 or S_2 . Suppose that $\mathbb{B}^3 \setminus D_w(S_2^{-1}) \subseteq D_w(S_1)$ and $\mathbb{B}^3 \setminus D_w(S_1^{-1}) \subseteq D_w(S_2)$. Then the semigroup T generated by $S_1 \cup S_2$ is semidiscrete. Furthermore, if S_1 and S_2 are inverse free, then so is T .*

PROOF. Observe that, for $g \in S_1 \setminus \{I\}$ and $h \in S_2 \setminus \{I\}$, we have

$$K(g(w), w)^\circ \subseteq \mathbb{B}^3 \setminus D_w(S_1) \subseteq D_w(S_2^{-1}) \subseteq K(w, h^{-1}(w)),$$

which implies that $K(g(w), w)^\circ \subseteq K(w, h^{-1}(w))^\circ$. Hence $h(K(g(w), w)^\circ) \subseteq K(h(w), w)^\circ$, and similarly $g(K(h(w), w)^\circ) \subseteq K(g(w), w)^\circ$. It follows that h maps $\mathbb{B}^3 \setminus D_w(S_1)$ into $\mathbb{B}^3 \setminus D_w(S_2)$, and g maps $\mathbb{B}^3 \setminus D_w(S_2)$ into $\mathbb{B}^3 \setminus D_w(S_1)$.

Next, observe that $g(w) \in \mathbb{B}^3 \setminus D_w(S_1)$ and $h(w) \in \mathbb{B}^3 \setminus D_w(S_2)$. Choose $f \in T \setminus \{I\}$. By writing f as a reduced word with letters from S_1 and S_2 , we see that $f(w)$ belongs to either $\mathbb{B}^3 \setminus D_w(S_1)$ or $\mathbb{B}^3 \setminus D_w(S_2)$. However, S_1 and S_2 are semidiscrete, so w is contained in the interior of both $D_w(S_1)$ and $D_w(S_2)$. It follows that T is semidiscrete.

Furthermore, if S_1 and S_2 are inverse free, then the same argument shows that no nontrivial reduced word with letters from S_1 and S_2 is equal to the identity element. Hence T is inverse free also. \square

Theorem 2.17 says that if S is countable, semidiscrete and inverse free, then $e_w(S)$ has nonempty interior if and only if $\Lambda^+(S) \neq \mathbb{S}^2$. This implies that if $\Lambda^+(S) \neq \mathbb{S}^2$ and $\Lambda^-(S) \neq \mathbb{S}^2$, then by Theorem 2.20 we can always choose another Möbius transformation f that does not lie in S , such that the semigroup $T = \langle \{f\} \cup S \rangle$ is semidiscrete and inverse free. The construction also shows that $\Lambda^+(T)$ is contained in the Euclidean closure of $(\mathbb{B}^3 \setminus D_w(S)) \cup K(f(w), w)$, and so we can choose f such that $\Lambda^+(T) \neq \mathbb{S}^2$.

Theorem 2.20 can be used as a tool for modifying known semidiscrete semigroups to generate new semidiscrete semigroups, as the following example illustrates. Let $S_1 = \langle f_0, f_1, f_2 \rangle$ and $S_2 = \langle g_0, g_1, g_2 \rangle$, where each generator is defined by the parameters in the table following Figure 2.5. Both semigroups are semidiscrete and inverse free. The Dirichlet regions of S_1 , S_1^{-1} , S_2 and S_2^{-1} with respect to the point $w = 0$ are shown in the figures below. The Dirichlet region of S_1 (the complement of the red region) and S_1^{-1} (the complement of the blue region) are shown in Figure 2.3. The Dirichlet region of S_2 (the complement of the red region) and S_2^{-1} (the complement of the blue region) are shown in Figure 2.4. Notice that the red region in Figure 2.3 does not meet the blue region in Figure 2.4, although it does meet the red region in Figure 2.4. Similarly the red region in Figure 2.4 does not meet the blue region in Figure 2.3, although it does meet the red region in Figure 2.3. Hence, by the procedure described above, the semigroup S generated by $S_1 \cup S_2$ is also semidiscrete and inverse free. The Dirichlet regions of S (the complement of the red region) and of S^{-1} (the complement of the blue region), are depicted in Figure 2.5.

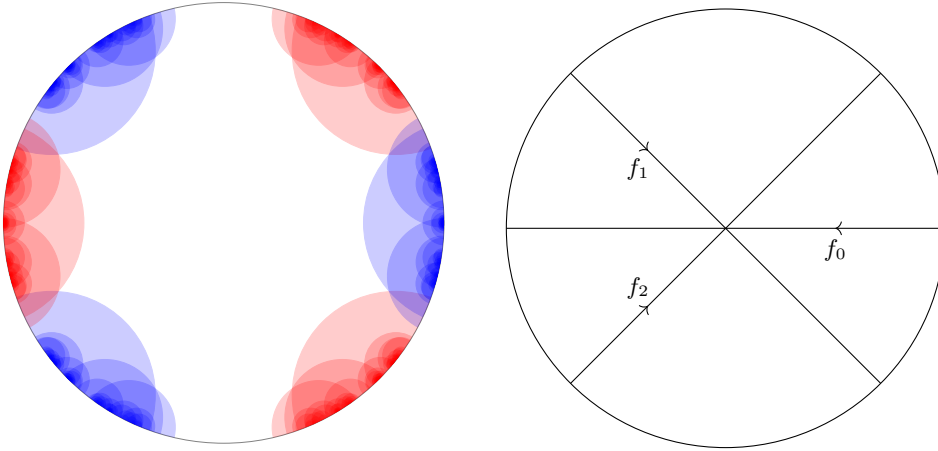


FIGURE 2.3. The generators of S_1 together with the Dirichlet regions of S_1 and S_1^{-1}

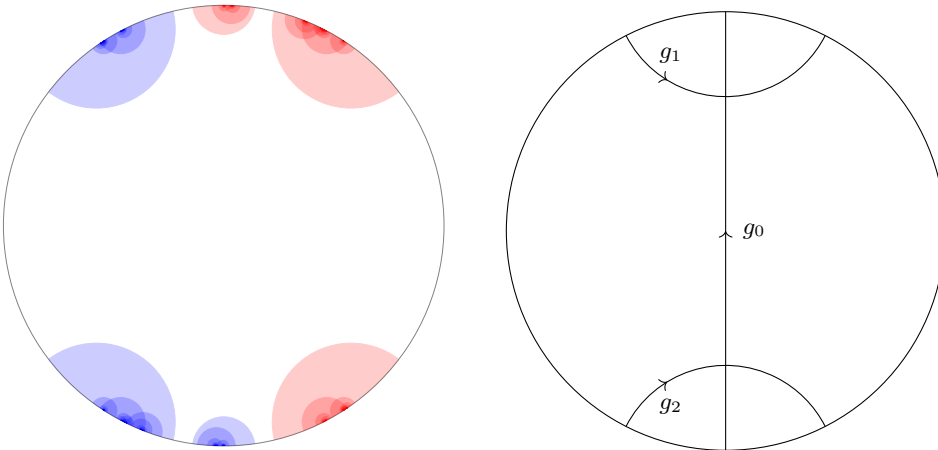


FIGURE 2.4. The generators of S_2 together with the Dirichlet regions of S_2 and S_2^{-1}

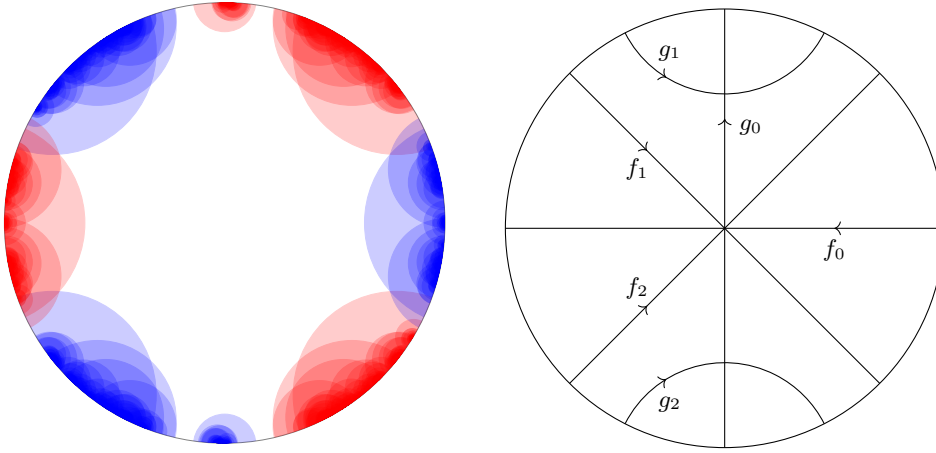


FIGURE 2.5. The generators of S together with the Dirichlet regions of S and S^{-1}

Transformation h	α_h	β_h	$\text{tr}^2(h)$
f_0	$e^{\pi i}$	$e^{0\pi i}$	22.05
f_1	$e^{0.25\pi i}$	$e^{1.25\pi i}$	22.05
f_2	$e^{1.75\pi i}$	$e^{0.75\pi i}$	22.05
g_0	$e^{0.5\pi i}$	$e^{1.5\pi i}$	202.005
g_1	$e^{0.35\pi i}$	$e^{0.65\pi i}$	$10.45\bar{3}$
g_2	$e^{1.35\pi i}$	$e^{1.65\pi i}$	$10.45\bar{3}$

CHAPTER 3

Semigroups that fix the unit circle

In this chapter we begin to present the main results of this thesis. Although many of the auxiliary results hold in all dimensions, most of the main results within this section apply only to subsemigroups of $\text{Aut}(\mathbb{D})$. Our arguments usually do not generalise to higher dimensions because Theorem 3.8, upon which other results rely, depends on the topological property that any connected subset of the ideal boundary consisting of at least two points contains a non-empty open subset. Nevertheless, some of our results *do* extend to subsemigroups of Möbius transformations that fix the unit circle. This means that we can include Möbius transformations that transpose the two connected components of $\overline{\mathbb{C}} \setminus \mathbb{S}^1$, as well as those that fix both components.

The group of Möbius transformations that fix the unit circle, which we denote $\text{Aut}(\mathbb{S}^1)$, is significant in the theory of continued fractions because, thinking in the upper half-plane model, although $z \mapsto z + 1$ fixes the upper half-plane, the map $z \mapsto 1/z$ transposes the upper and lower half-planes. The unit circle in the complex plane identifies with the equator in the unit ball model. This means that, in the unit ball model, since any element of $\text{Aut}(\mathbb{S}^1)$ fixes the equator it must also fix the Euclidean disc $\{(x, y, z) \in \mathbb{B}^3 \mid z = 0\}$ contained in \mathbb{B}^3 whose boundary is the equator. Indeed, $\text{Aut}(\mathbb{S}^1)$ is exactly the set of isometries (both orientation-preserving and orientation-reversing) of $\{(x, y, z) \in \mathbb{B}^3 \mid z = 0\}$ endowed with the hyperbolic metric inherited from \mathbb{B}^3 . Each element of $\text{Aut}(\mathbb{S}^1)$ either fixes the southern hemisphere (which corresponds to the unit disc in the complex plane) or transposes the northern and southern hemispheres. In this chapter we are concerned with subsemigroups of $\text{Aut}(\mathbb{D})$, however, by working in the ball model and considering the action of $\text{Aut}(\mathbb{S}^1)$ on $\{(x, y, z) \in \mathbb{B}^3 \mid z = 0\}$, our arguments that do not require this action to be orientation-preserving (for example Theorem 3.1) also hold for subsemigroups

of $\text{Aut}(\mathbb{S}^1)$.

We now survey the topics discussed within this chapter. In Section 2 we develop an understanding of composition sequences generated by finitely-generated semidiscrete subsemigroups of $\text{Aut}(\mathbb{D})$. In particular we prove the following theorem.

THEOREM 3.1. *Let $S = \langle \mathcal{F} \rangle$ be a finitely-generated subsemigroup of $\text{Aut}(\mathbb{D})$. The following statements are equivalent:*

- (i) *every composition sequence generated by S is an escaping sequence;*
- (ii) *every composition sequence generated by S converges ideally;*
- (iii) *S is semidiscrete and inverse free.*
- (iv) *every composition sequence generated by \mathcal{F} is an escaping sequence;*
- (v) *every composition sequence generated by \mathcal{F} converges ideally;*

This tells us that for any semidiscrete subsemigroup of $\text{Aut}(\mathbb{D})$, every composition sequence converges ideally if and only if the group part is empty. We also describe how composition sequences can behave when S has a nonempty group part.

In Section 3 we give an algorithm for deciding whether or not a two-generator subsemigroup of $\text{Aut}(\mathbb{D})$ is semidiscrete, semidiscrete and inverse free, or neither. Our methods also give a classification of semidiscrete two-generator semigroups (Theorem 3.20). Our strategy is similar to the geometric approach for classifying discrete two-generator groups initiated by Matelski [31] and others, and completed by Gilman [17]. However, our task is much simpler than that of classifying discrete two-generator groups. This is because by far the most difficult case in the groups classification arises when the two generators are loxodromic maps with intersecting axes. In contrast, the semigroup generated by two such maps is easily seen to be semidiscrete.

It is well known that any group of Möbius transformations is either elementary, Fuchsian, or dense in $\text{Aut}(\mathbb{D})$. Section 4 is concerned with proving a counterpart of this result for subsemigroups of $\text{Aut}(\mathbb{D})$. It has a similar statement, but with an additional category. Let J be a nontrivial closed interval of \mathbb{S}^1 . Throughout this chapter whenever we say ‘closed

interval' we shall mean a nontrivial closed interval (that is nonempty, and not a single point or the whole unit circle) contained in the unit circle. We define $\text{Möb}(J)$ to be the collection of Möbius transformations in $\text{Aut}(\mathbb{D})$ that map J within itself. Clearly $\text{Möb}(J)$ is a semigroup, which is not semidiscrete. We shall prove the following result.

THEOREM 3.2. *Let S be a subsemigroup of $\text{Aut}(\mathbb{D})$. Then S is either*

- (i) *elementary,*
- (ii) *semidiscrete,*
- (iii) *contained in $\text{Möb}(J)$, for some nontrivial closed interval J , or*
- (iv) *dense in $\text{Aut}(\mathbb{D})$.*

If S is finitely-generated, then most semigroups that are contained in $\text{Möb}(J)$ for some interval J (of type (iii)) are also semidiscrete (of type (ii)). There are some exceptions to this, however; for example, (now using the upper half-plane model) the semigroup contained in $\text{Aut}(\mathbb{H}^2)$ and generated by the maps $z \mapsto \sqrt{2}z$, $z \mapsto \frac{1}{2}z$ and $z \mapsto z + 1$. This semigroup is contained in $\text{Möb}([0, +\infty])$, but it is not elementary, semidiscrete or dense in $\text{Aut}(\mathbb{H}^2)$. In Section 5 we will classify the small collection of finitely-generated semigroups that are not of types (i), (ii) or (iv). We will see that S does not lie in this special collection of semigroups if no two members of a generating set for S are loxodromic with the same attracting and repelling fixed points. This version of Theorem 3.2 for finitely-generated semigroups represents a significant generalisation of a result of Bárány, Beardon and Carne [2, Theorem 3], who proved that any semigroup generated by two non-commuting transformations, one of which is elliptic of infinite order (that is, an elliptic map g for which $g^n \neq I$ for any $n \in \mathbb{N}$), is dense in $\text{Aut}(\mathbb{D})$.

In Section 5 we prove another theorem which has a familiar counterpart in the theory of Fuchsian groups. This counterpart theorem says that a nonelementary group of Möbius transformations is discrete if and only if each two-generator subgroup of the group is discrete. Our theorem only applies to finitely-generated subsemigroups of $\text{Aut}(\mathbb{D})$. Unfortunately, there is also a bothersome class of semigroups that we must treat as exceptional cases, which we now describe.

Definition Let S be a semigroup that lies in $\text{Möb}(J)$, for some closed interval J . Suppose that the collection of elements of S that fix J as a set forms a nontrivial discrete group. Suppose also that one of the other members of S (outside this discrete group) fixes one of the end points of J . In these circumstances we say that S is an *exceptional* semigroup; otherwise it is a *nonexceptional* semigroup.

These terms should be treated in a similar way to how we treat the terms *elementary* and *nonelementary* in the theory of groups or semigroups; that is, exceptional semigroups, like elementary groups or semigroups, are easy to handle, but do not obey all the laws satisfied by the more typical cases. We emphasise that it is easy to tell whether a finitely-generated semigroup is exceptional by examining its generating set, as we explain later. The main result of Section 5 is the following.

THEOREM 3.3. *Suppose S is a finitely-generated nonexceptional semigroup S contained in $\text{Aut}(\mathbb{D})$, that is not elementary of infinite type. Then S is semidiscrete if and only if every two-generator semigroup contained in S is semidiscrete.*

We use the results on exceptional semigroups developed in this section to give a version of Theorem 3.2 for finitely-generated semigroups. This is Corollary 3.30 and can be regarded as a (coarse) classification of finitely-generated subsemigroups of $\text{Aut}(\mathbb{D})$.

Section 6 covers our final theorem, on limit sets of semigroups. There appears to be a relationship between the size of the intersection $\Lambda^+(S) \cap \Lambda^-(S)$ and the size of the group $S \cap S^{-1}$. For example, if $\Lambda^+(S) \cap \Lambda^-(S)$ is finite, then the group $S \cap S^{-1}$ is either empty or elementary (because if it is not elementary, then its limit set, which is contained in both $\Lambda^+(S)$ and $\Lambda^-(S)$, is perfect). The main result of this section is about when the intersection $\Lambda^+(S) \cap \Lambda^-(S)$ is large.

THEOREM 3.4. *Let S be a finitely-generated nonelementary subsemigroup of $\text{Aut}(\mathbb{D})$. Then $\Lambda^+(S) = \Lambda^-(S)$ if and only if S is a group.*

In fact, we prove that if $\Lambda^-(S) \subseteq \Lambda^+(S)$, then S is a group.

Before proceeding, we take a more detailed look at elementary subsemigroups of $\text{Aut}(\mathbb{D})$, or equivalently elementary subsemigroups of $\text{Aut}(\mathbb{H}^2)$, and in particular give the promised version of Theorem 1.7 from Chapter 1 which has been specialised to semigroups contained in $\text{Aut}(\mathbb{H}^2)$.

1. A closer look at elementary subsemigroups of $\text{Aut}(\mathbb{H}^2)$

THEOREM 3.5. *Let S be an elementary semigroup of finite type contained in $\text{Aut}(\mathbb{H}^2)$. Then S is conjugate in $\text{Aut}(\mathbb{H}^2)$ to a semigroup that is contained in one of the following sets, depending on $|\Lambda^-|$ and $|\Lambda^+|$:*

- (i) $|\Lambda^-| = |\Lambda^+| = 0$: $\{z \mapsto (az - b)/(bz + a) \mid a^2 + b^2 = 1\}$;
- (ii) $|\Lambda^-| = |\Lambda^+| = 1$: $\{z \mapsto z + a \mid a \in \mathbb{R}\}$ or $\{z \mapsto \lambda z \mid \lambda > 0\}$;
- (iii) $|\Lambda^-| = |\Lambda^+| = 2$: $\{z \mapsto \lambda z, z \mapsto -\lambda/z \mid \lambda > 0\}$;
- (iv) (a) $|\Lambda^-| = 1, |\Lambda^+| = \infty$: $\{z \mapsto az + b \mid a \leq 1, b \in \mathbb{R}\}$;
- (b) $|\Lambda^-| = \infty, |\Lambda^+| = 1$: $\{z \mapsto az + b \mid a \geq 1, b \in \mathbb{R}\}$.

The proof of this theorem is elementary, and similar to the proof of the classification of elementary semigroups of Möbius transformations given in [15, Theorem 2.11]. Accordingly, we only sketch a justification of the theorem.

PROOF OF THEOREM 3.5. Suppose that one of Λ^- and Λ^+ is empty. Then S contains only elliptic transformations, which each fix a common point. If, by conjugation, we assume that the fixed point is i , then S is contained in $\{z \mapsto (az - b)/(bz + a) \mid a^2 + b^2 = 1\}$. Suppose now that $|\Lambda^+| = 2$. After conjugating, we can assume that $\Lambda^+ = \{0, \infty\}$. Since Λ^+ is forward invariant under S , every element of S must either fix 0 and ∞ , or interchange them. The elements of $\text{Aut}(\mathbb{D})$ that fix both 0 and ∞ have the form $z \mapsto \lambda z$, $\lambda > 0$, and the elements of $\text{Aut}(\mathbb{D})$ that interchange 0 and ∞ have the form $z \mapsto -\lambda/z$, $\lambda > 0$. The former type are loxodromic (unless $\lambda = 1$) and the latter type are elliptic. Since $\infty \in \Lambda^+$, it must be an attracting fixed point of some map $z \mapsto \lambda z$, $\lambda > 1$. Hence $0 \in \Lambda^-$, and similarly $\infty \in \Lambda^-$. Therefore $\Lambda^- = \{0, \infty\}$.

We are left to consider the cases when one of $|\Lambda^-|$ or $|\Lambda^+|$ has order 1. If both have order 1, then we are in case (ii); we omit the details. If only one has order 1, say $|\Lambda^-| = 1$ (the other case is similar), then by conjugation we can assume that $\Lambda^- = \{\infty\}$. The elements

of $\text{Aut}(\mathbb{D})$ that fix ∞ have the form $z \mapsto az + b$, where $a > 0$ and $b \in \mathbb{R}$. If $a > 1$, then this map is loxodromic and has a *repelling* fixed point at $b/(1 - a)$. As all repelling fixed points of loxodromic elements of S belong to Λ^- , we see that S is of type (iv)(a). \square

Later on we will require a more precise understanding of those finitely-generated semigroups of type (ii). The following lemma satisfies our needs: it classifies those semigroups that are contained in the group $\{z \mapsto z + a \mid a \in \mathbb{R}\}$. This task is the same as classifying additive semigroups of real numbers, and so some form of this lemma almost certainly features in the literature already.

LEMMA 3.6. *Let S be a semigroup generated by maps $z \mapsto z + a_i$, $i = 1, \dots, n$. Then exactly one of the following statements is true.*

- (a) *Either $a_i \geq 0$ for $i = 1, \dots, n$ or $a_i \leq 0$ for $i = 1, \dots, n$, in which case S is semidiscrete.*
- (b) *For every pair of indices i and j , we have $a_i/a_j \in \mathbb{Q}$, provided $a_j \neq 0$, and one of these quotients a_i/a_j is negative. In this case, S is a discrete group.*
- (c) *Otherwise, there are two maps $z \mapsto z + a_i$ and $z \mapsto z + a_j$ that generate a dense subset of $\{z \mapsto z + a \mid a \in \mathbb{R}\}$.*

PROOF. In case (a), it is clear that S is semidiscrete.

In case (b), choose any of the numbers a_i (other than 0), and let a_j be of opposite sign to a_i . Then $ma_i = -na_j$ for some coprime positive integers m and n . Choose positive integers u and v such that $vn - um = \pm 1$. Let $\alpha = a_i/n = -a_j/m$. Then $va_i + ua_j = \pm\alpha$. It follows that S contains one of the maps $z \mapsto z + \alpha$ or $z \mapsto z - \alpha$. But S also contains $z \mapsto z + n\alpha$ and $z \mapsto z - m\alpha$. So S must contain *both* $z \mapsto z + \alpha$ and $z \mapsto z - \alpha$, and in particular it must contain $z \mapsto z - a_i$. Therefore S is a group, and one can use a short, standard argument from the theory of discrete groups to prove that S is discrete.

In case (c), we can choose a pair of nonzero numbers a_i and a_j such that $a_i/a_j \notin \mathbb{Q}$. We can assume that a_i and a_j have opposite signs: if at first they do not, then replace one of them by another number a_k of the opposite sign (if you are careful about which of a_i and a_j you replace, then you can be sure that the quotient of the two numbers you end up with is irrational). Now let u_n/v_n be a sequence of rational numbers, where $u_n, v_n > 0$,

that satisfies

$$\left| \frac{a_i}{a_j} + \frac{u_n}{v_n} \right| < \frac{1}{2v_n^2}.$$

Then $|v_n a_i + u_n a_j| \rightarrow 0$ as $n \rightarrow \infty$, and we know that $v_n a_i + u_n a_j \neq 0$. From this we see that the maps $z \mapsto z + v_n a_i + u_n a_j$ accumulate at the identity, so $\langle z \mapsto z + a_i, z \mapsto z + a_j \rangle$ is not semidiscrete. \square

The group $\{z \mapsto z + a \mid a \in \mathbb{R}\}$ considered in the lemma is the group of orientation-preserving Euclidean isometries of \mathbb{R} . The other group in Theorem 3.5(ii) is $\{z \mapsto \lambda z \mid \lambda > 0\}$, which is the group of orientation-preserving isometries of one-dimensional hyperbolic space \mathbb{I} , which in this case is modelled by the positive imaginary axis. The metric spaces \mathbb{R} and \mathbb{I} are isometric using the transformation $x \mapsto e^x$. It follows that the classification of finitely-generated semigroups of $\{z \mapsto \lambda z \mid \lambda > 0\}$ is much the same as the classification of Lemma 3.6; if S is generated by maps $z \mapsto \lambda_i z$, $i = 1, \dots, n$, then we can define $a_i = \log \lambda_i$ and cases (a)–(c) remain valid, with minor adjustments.

We have now examined class (ii) of Theorem 3.5 in detail. Let us briefly survey the other classes in turn, without proofs. Throughout our analysis we assume that S is a finitely-generated semigroup that is contained in a set G , where G is specified for each class.

Class (i). Let $G = \{z \mapsto (az - b)/(bz + a) \mid a^2 + b^2 = 1\}$. Then S is either dense in G or else it is a finite cyclic subgroup of G .

Class (iii). Let $G = \{z \mapsto \lambda z, z \mapsto -\lambda/z \mid \lambda > 0\}$. Assume that S contains a map of the form $z \mapsto -\lambda/z$ (otherwise S has been dealt with already in class (ii)). Then S is a group, so either it is dense in G or else it is conjugate in $\text{Aut}(\mathbb{B}^3)$ to a discrete group generated by $z \mapsto -1/z$ and $z \mapsto \mu z$, where $\mu > 1$.

Class (iv)(a). Let $G = \{z \mapsto az + b \mid a \leq 1, b \in \mathbb{R}\}$. Define $S_1 = \{f(z) = z + b \mid f \in S\}$. Then S_1 is finitely-generated, so either it is empty, or else it is one of the types (a)–(c) determined in the first part of the analysis of class (ii), above. If S_1 is of type (c), then S is not semidiscrete, but it is not dense in G . If S_1 is empty or of types (a) or (b), then S is a semidiscrete semigroup.

Class (iv)(b). This is similar to the previous case (switch between S and S^{-1} to move from one class to the other).

2. Proof of Theorem 3.1

In this section it is convenient to use the upper half-plane model of the hyperbolic plane, whose ideal boundary is the extended real line $\overline{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$. Before proving Theorem 3.1, we need a basic lemma.

LEMMA 3.7. *Suppose that S is a semigroup generated by a finite set \mathcal{F} of Möbius transformations, and suppose that F_n is an escaping composition sequence generated by \mathcal{F} . Let $L = \Lambda_c^+(F_n)$. Then for any positive integer m , $F_m^{-1}(L) \subseteq \Lambda_c^+(S)$.*

PROOF. Let $p \in L$. Choose $\zeta \in \mathbb{H}^2$. Then there is a sequence of positive integers n_1, n_2, \dots such that the sequence $F_{n_i}(\zeta)$ converges conically to p . Hence the sequence $F_m^{-1}F_{n_i}(\zeta)$ converges conically to $F_m^{-1}(p)$. Now, if $n_i > m$, then $F_m^{-1}F_{n_i} \in S$, and so $F_m^{-1}(p) \in \Lambda_c^+(S)$, as required. \square

The lemma remains true if we replace the forward *conical limit sets* with forward *limit sets*; however, we shall not need this alternative statement.

THEOREM 3.8. *Suppose that S is a semigroup generated by a finite set \mathcal{F} contained in $\text{Aut}(\mathbb{D})$. Suppose also that there is an escaping composition sequence F_n generated by \mathcal{F} that does not converge ideally. Then $\Lambda_c^+(S) = \mathbb{S}^1$, unless $\Lambda^-(S) = \{q\}$ for some point q , in which case $\Lambda_c^+(S)$ is either equal to \mathbb{S}^1 or to $\mathbb{S}^1 \setminus \{q\}$.*

PROOF. Working in the upper half-plane model, choose any point ζ in \mathbb{H}^2 . Let $k = \max\{\rho(\zeta, f(\zeta)) \mid f \in \mathcal{F}\}$. We are given that the sequence $F_n(\zeta)$ accumulates, in the chordal metric, at two distinct points a and b in $\overline{\mathbb{R}}$. Suppose there is a point d other than a and b that is *not* a forward conical limit point of F_n . By conjugating F_n if need be we can assume that $d = \infty$ and $a < b$. Let γ_c be the hyperbolic geodesic with one end point c inside (a, b) and the other at ∞ . Let $\Gamma_c = \{z \in \mathbb{H}^2 \mid \rho(z, \gamma_c) < k\}$. Notice that

$$\rho(F_{n-1}(\zeta), F_n(\zeta)) = \rho(\zeta, f_n(\zeta)) \leq k.$$

It follows that infinitely many terms from the sequence $F_n(\zeta)$ lie in Γ_c . This infinite set inside Γ_c cannot accumulate at ∞ because ∞ is not a forward conical limit point of F_n , so it must accumulate at c . Hence c is a forward conical limit point of F_n .

We have shown that $(a, b) \subseteq \Lambda_c^+(F_n)$. Choose a point y in (a, b) . Since $\Lambda_c^-(F_n^{-1}) = \Lambda_c^+(F_n)$, we can apply Lemma 1.3 to deduce the existence of a sequence of positive integers n_1, n_2, \dots and two distinct points p and q in $\overline{\mathbb{R}}$ such that $F_{n_i}^{-1}(y) \rightarrow p$ and $F_{n_i}^{-1}(z) \rightarrow q$ for $z \neq y$. It follows that any point u in $\overline{\mathbb{R}} \setminus \{q\}$ is contained in $F_{n_i}^{-1}(a, b)$, providing n_i is sufficiently large. Lemma 3.7 tells us that $F_{n_i}^{-1}(a, b) \subseteq \Lambda_c^+(S)$, so we see that $\overline{\mathbb{R}} \setminus \{q\} \subseteq \Lambda_c^+(S)$.

Finally, suppose that $\Lambda^-(S) \neq \{q\}$. If there exists an element f of S that does not fix q , then $q = f(v)$ for some point v in $\overline{\mathbb{R}} \setminus \{v\}$. As $\Lambda_c^+(S)$ is forward invariant under S , we see that it is the whole of $\overline{\mathbb{R}}$. Otherwise if every element of S fixes q , then S contains a loxodromic element with attracting fixed point q , in which case $q \in \Lambda_c^+(S)$ and so again we have $q \in \Lambda_c^+(S) = \overline{\mathbb{R}}$. \square

The exceptional case in which $\Lambda^-(S) = \{q\}$ and $\Lambda_c^+(S) = \overline{\mathbb{R}} \setminus \{q\}$ certainly can arise. For example, let $S = \langle z \mapsto \frac{1}{2}z, z \mapsto z+1, z \mapsto z-1 \rangle$. Clearly $\Lambda^-(S) = \{\infty\}$, as all the generators of S fix ∞ , and ∞ is a repelling fixed point of $z \mapsto \frac{1}{2}z$. One can easily construct a composition sequence generated by S that is an escaping sequence that does not converge ideally, and one can check that $\Lambda_c^+(S)$ contains \mathbb{R} . However, $\infty \notin \Lambda_c^+(S)$ because all elements of S have the form $z \mapsto az + b$, where $a \leq 1$ (and so the S -orbit of a point in \mathbb{H}^2 cannot accumulate conically to ∞).

A Kleinian group G is said to be *cocompact* if its quotient space \mathbb{B}^3/G is compact, or equivalently, if G has a compact fundamental region. We can now prove the main result of this section, from which we will deduce Theorem 3.1.

THEOREM 3.9. *Let $S = \langle \mathcal{F} \rangle$ be a finitely-generated semidiscrete subsemigroup of $\text{Aut}(\mathbb{D})$ such that $|\Lambda^-(S)| \neq 1$. Then the following statements are equivalent:*

- (i) *every composition sequence generated by \mathcal{F} that is an escaping sequence converges ideally;*
- (ii) *S is not a cocompact Fuchsian group.*

PROOF. Suppose first that S is a cocompact Fuchsian group. Then it has a finite-sided compact fundamental polygon D and the images of D under S tessellate \mathbb{D} . Choose z_0 in the interior of D and suppose $f_1(D), f_2(D), \dots, f_k(D)$ are the neighbouring tiles of D . Let

$D_0 = D, D_1, D_2, \dots$ be a sequence of tiles such that D_i is adjacent to D_{i+1} for each i . Let z_i be the point of $S(z_0)$ that lies in D_i . We can choose the sequence D_i so that the sequence z_i accumulates at two points of the unit circle \mathbb{S}^1 , but does not accumulate in \mathbb{D} . For each $n \geq 0$ there exists a unique element $g_n \in S$ such that $g_n(D) = D_n$. The neighbours of D_n are $g_n(f_i(D))$, and one of these is $D_{n+1} = g_{n+1}(D)$, and so $g_{n+1} = g_n f_{i_{n+1}}$ for some $i_{n+1} \in \{1, \dots, k\}$. Therefore $g_n = f_{i_1} \cdots f_{i_n}$ defines a composition sequence generated by the finite set $\mathcal{F} = \{f_1, \dots, f_k\}$. Since $g_n(z_0) = z_n$, the sequence is escaping but does not converge ideally.

Suppose, conversely, that there is a composition sequence generated by \mathcal{F} that is an escaping sequence but does not converge ideally. By Theorem 3.8, we see that $\Lambda_c^+(S) = \mathbb{S}^1$. Hence $\Lambda_h^+ = \mathbb{S}^1$ and so Corollary 2.15 then tells us that S is a group with a compact Dirichlet region: it is a cocompact Fuchsian group. \square

Suppose \mathcal{F} is a finite subset of $\text{Aut}(\mathbb{D})$ that generates a semidiscrete semigroup S such that $|\Lambda^-(S)| \neq 1$. Further suppose that S is not a cocompact Fuchsian group. If $F_n = f_1 \cdots f_n$ is a composition sequence generated by \mathcal{F} , then there are two cases: either F_n escapes, or not. If F_n escapes, then it converges ideally by Theorem 3.9. Otherwise F_n does not escape, and for all large enough n each f_n belongs to the group part of S . To see this, we have by Lemma 2.6 that along some subsequence n_k , the sequence F_{n_k} converges to a Möbius transformation. Hence, as in the proof of Theorem 2.7, $F_{n_k}^{-1} F_{n_{k+1}}$ converges to the identity. Since S is semidiscrete, this means that $F_{n_k}^{-1} F_{n_{k+1}} = f_{n_k+1} \cdots f_{n_{k+1}}$ is equal to the identity for all large enough k . It follows that for all large enough n , each f_n lies in the group part of S .

Theorem 3.9 fails if we remove the hypothesis $|\Lambda^-(S)| \neq 1$; we have seen an example of this already, just before the statement of the theorem. Let us now use this theorem to prove Theorem 3.1.

PROOF OF THEOREM 3.1. The equivalence of (i) and (iii) was established already in Theorem 2.7. That (ii) implies (i) is immediate. The most substantial part of Theorem 3.1 is the implication of (ii) from (iii).

Suppose then that S is an inverse free semidiscrete semigroup, generated by a finite set \mathcal{F} . We must prove that every composition sequence generated by S converges ideally. By the equivalence of (i) and (iii), we know that every composition sequence generated by S is an escaping sequence. If we regard any composition sequence generated by S as a subsequence F_{n_k} of a composition sequence F_n generated by \mathcal{F} , then Theorem 3.9 tells us that, provided $|\Lambda^-(S)| \neq 1$, F_n converges ideally, and so F_{n_k} also converges ideally. In the special case where $|\Lambda^-(S)| = 1$ we can assume, by conjugating, that $\Lambda^-(S) = \{\infty\}$. Now by examining Class (iv)(a) in Theorem 3.5, we see that there is a real number a , such that all elements of S either map the interval $(-\infty, a)$ within itself, or map the interval $(a, +\infty)$ within itself. Either way Theorem 3.8 tells us that every composition sequence generated by S that is an escaping sequence must converge ideally.

We see that (i) and (iv) are equivalent by Theorem 2.7. To show that (v) and (ii) are equivalent we argue just as in the proof of Theorem 2.7: any composition sequence generated by S is a subsequence of some composition sequence generated by \mathcal{F} , and so (v) implies (ii). Conversely since $\mathcal{F} \subseteq S$ it follows that (ii) implies (v). \square

3. Two-generator semigroups

In this section we classify the semidiscrete two-generator subsemigroups of $\text{Aut}(\mathbb{D})$, and we determine which of them are Fuchsian groups and which are inverse free semigroups. This section can be thought of as a version of Gilman and Maskit's paper [18] for semidiscrete semigroups. There, the authors give an algorithm that determines whether or not a two-generator subgroup of $\text{Aut}(\mathbb{D})$ is discrete, which omits the more complicated case of two loxodromic generators with intersecting axes, which Gilman deals with in [17]. This section relates only to subsemigroups of $\text{Aut}(\mathbb{D})$, and so throughout all Möbius transformations are understood to be of this type. The task of classification is simpler if we handle elementary two-generator semigroups separately to the main classification of two-generator semigroups.

THEOREM 3.10. *The semigroup $S = \langle g, h \rangle$ generated by two nontrivial Möbius transformations g and h in $\text{Aut}(\mathbb{D})$ is elementary if and only if*

- (i) *g and h have a common fixed point in $\overline{\mathbb{D}}$; or*

- (ii) g and h are both elliptic of order 2; or
- (iii) one of g and h is loxodromic and the other is elliptic of order 2, and the fixed point of the elliptic generator lies on the axis of the loxodromic generator.

PROOF. To begin note that if $f \in \text{Aut}(\mathbb{D})$ fixes $w \in \overline{\mathbb{C}}$ then it also fixes $1/\overline{w}$. We will work through the cases listed in Theorem 1.5.

If S fixes a point x in \mathbb{H}^3 then S consists of elliptic transformations. The axis of any elliptic transformation in $\text{Aut}(\mathbb{D})$ has ideal endpoints w and $1/\overline{w}$ for some w in \mathbb{D} . So S consists of elliptic transformations that fix the unique point w in \mathbb{D} such that the geodesic that ends at w and $1/\overline{w}$ passes through x , and we are in case (i).

Suppose S is elementary but does not fix any point in \mathbb{H}^3 . Then either S fixes a single point w in $\overline{\mathbb{C}}$ or a pair of points v, w in $\overline{\mathbb{C}}$. In the first case, w must lie on the unit circle, since $1/\overline{w}$ is also fixed by S , so again we are in case (i). In the second case, we cannot have $v = 1/\overline{w}$ because then v would be in \mathbb{D} and $w \notin \mathbb{D}$ or vice versa, and S would fix pointwise the geodesic joining v and w in \mathbb{H}^3 . We must therefore have $v = 1/\overline{v}$ and $w = 1/\overline{w}$, so v and w are on the unit circle. Either g and h both fix v and w , so we are in case (i), or they both exchange v and w , so we are in case (ii), or one fixes them and the other exchanges them, so we are in case (iii). \square

There are examples of semigroups of both finite and infinite type in case (i), however every semigroup in case (ii) and case (iii) are of finite type. Using Theorem 3.10 it is straight forward to decide which two-generator elementary semigroups of finite type are semidiscrete and which are inverse free. The only two-generator elementary semigroups that are of infinite type are described in the following proposition, and are all semidiscrete and inverse free.

THEOREM 3.11. *Suppose g and h are two nontrivial Möbius transformations in $\text{Aut}(\mathbb{D})$ that generate an elementary semigroup of infinite type. Then S is semidiscrete and inverse free.*

PROOF. Suppose g and h are nontrivial Möbius transformations in $\text{Aut}(\mathbb{D})$ and that $S = \langle g, h \rangle$ is an elementary semigroup of infinite type. By the comments following Theorem 1.5, both g and h fix some point w in $\overline{\mathbb{C}}$. If w is not on the unit circle, then S fixes a

point in the unit disc (either w or $1/\overline{w}$), and so both limit sets are empty, contrary to our assumption that S is of infinite type. Hence w lies on the unit circle and both generators are not elliptic transformations. If at least one of g or h is a parabolic transformation, then it is easy to see that at least one of $\Lambda^+(S)$ or $\Lambda^-(S)$ is equal to $\{w\}$. Hence both g and h are loxodromic maps. If $w = \alpha_g = \alpha_h$ or $w = \beta_g = \beta_h$, then again, one of the limit sets of S is equal to $\{w\}$. If the set of fixed points of g is equal to the set of fixed points of h , then both $\Lambda^+(S)$ and $\Lambda^-(S)$ are contained in this set of fixed points. Since S is of infinite type, it follows that either S or S^{-1} is conjugate in $\text{Aut}(\mathbb{B}^3)$ to the configuration shown in Figure 3.1.

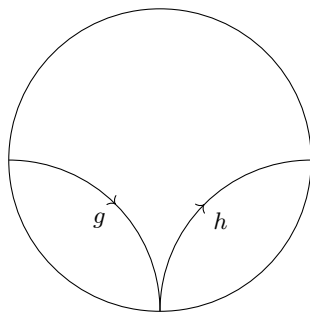


FIGURE 3.1. The axes of g and h

It now suffices to show that if the generators of S are configured as in Figure 3.1, then S is semidiscrete and inverse free. Choose a point z in the component of $\mathbb{S}^1 \setminus \{\beta_g, \alpha_h\}$ that does not contain the common fixed point $\alpha_g = \beta_h$. Let J be the closed interval bounded by z and the common fixed point $\alpha_g = \beta_h$ that contains α_h . Both g and h map J strictly within itself, and so by Theorem 2.5, S is semidiscrete and inverse free. \square

Now we have classified all two-generator semigroups in the cases where the generators have fixed points in common. We now proceed to consider two-generator semigroups for which any point fixed by one generator is not fixed by the other. We first consider the case where at least one generator is elliptic. Clearly S cannot be both semidiscrete and inverse free. However, if exactly one generator is elliptic and of finite order, and the composition of the two generators is not elliptic, then we have an example of a two-generator semidiscrete semigroup whose group part and inverse free part are both nonempty (see case (b) in

Theorem 3.12). In fact, examples of this type are the only two-generator semidiscrete semigroups whose group part and inverse free part are both nonempty. Throughout the rest of this section it is convenient to denote the group part of S , that is $S \cap S^{-1}$, by G .

THEOREM 3.12. *Suppose $S = \langle g, h \rangle$ where g is not elliptic and h is elliptic. Then exactly one of the following is true:*

- (a) h has infinite order and S is not semidiscrete;
- (b) h has finite order, gh^k is not elliptic for any $k = 0, \dots, \text{order}(h) - 1$, and S is semidiscrete;
- (c) h has finite order, and for some $k \in \{0, \dots, \text{order}(h) - 1\}$ the element gh^k is an elliptic transformation of infinite order, and S is not semidiscrete;
- (d) h has finite order, and for some $k \in \{0, \dots, \text{order}(h) - 1\}$ the element gh^k is an elliptic transformation of finite order, and S is the group generated by g and h .

Moreover any two-generator semidiscrete semigroup whose group part and inverse free part are both nonempty must arise from case (b).

PROOF. Suppose h is elliptic and g is either loxodromic or parabolic. If h has infinite order, then S is not semidiscrete and we have case (a). Now suppose h is of finite order, and by conjugating if necessary, we can suppose h has fixed point 0. Let m be the order of h . We can replace h with h^d , where d and m are coprime, without affecting $\langle g, h \rangle$. Therefore we can assume that the angle of rotation of h is $2\pi/m$.

If g is loxodromic, choose ℓ_γ to be the unique geodesic that is perpendicular to the axis of g and passes through 0. Let ℓ_α be the unique geodesic that is perpendicular to the axis of g and such that $g = \alpha\gamma$, where α and γ denote reflection in ℓ_α and ℓ_γ respectively.

If g is parabolic, choose ℓ_γ to be the unique geodesic that passes through 0 and lands at the fixed point of g . Let ℓ_α be the unique geodesic that lands at the fixed point of g and such that $g = \alpha\gamma$, where again α and γ denote reflection in ℓ_α and ℓ_γ respectively.

Now let us choose β to be the (unique) reflection such that $h = \gamma\beta$. Let ℓ_β denote the line of reflection of β , and note that the angle between ℓ_β and ℓ_γ is π/m . We define ℓ_{β_k} to be

the geodesic line passing through 0 obtained by rotating ℓ_γ by an angle of $k\pi/m$, and let β_k denote reflection in ℓ_{β_k} . Note that $\beta_0 = \gamma$, $\beta_1 = \beta$ and $h^k = \beta_k\beta_0$.

There are now two possibilities, either ℓ_α cuts ℓ_{β_k} in \mathbb{D} for some $k = 0, \dots, m-1$, or otherwise. In the latter case, for some k the line ℓ_α is contained in a sector of angle π/m subtended between ℓ_{β_k} and $\ell_{\beta_{k+1}}$. Let $\Gamma = \langle \beta_k, \beta_{k+1}, \alpha \rangle$ and let q denote whichever of $k, k+1$ is even. Since $\beta_0 = h^{-q/2}\beta_q h^{q/2}$ and $h = \beta_{k+1}\beta_k \in \Gamma$, it follows that $g = \alpha\beta_0$ must lie in Γ . Hence we have $S \subseteq \Gamma$. By Poincaré's Polygon Theorem 3.15, the shaded region bounded by ℓ_{β_k} , $\ell_{\beta_{k+1}}$ and ℓ_α shown in Figure 3.2 is a fundamental region for Γ . Since Γ is discrete S is semidiscrete, and so we have case (b).

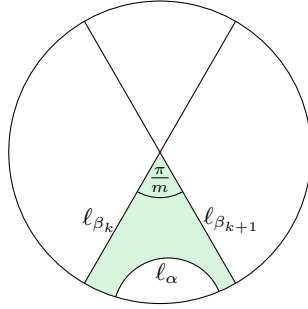


FIGURE 3.2. Fundamental region for Γ

Now let Y denote the closed arc contained in \mathbb{S}^1 whose end points are end points of ℓ_{β_k} and $\ell_{\beta_{k+1}}$, and which contains the end points of ℓ_α . Let $X = Y \cup h(Y) \cup \dots \cup h^{m-1}(Y)$. Then X is a nontrivial closed subset of \mathbb{S}^1 that satisfies $h(X) = X$ and, since $g = \alpha\beta_0$, satisfies $g(X) \subseteq Y$. It follows that $\langle g, h \rangle$ is not equal to the group generated as a group by g and h .

Earlier in the proof we assumed that ℓ_α did not meet ℓ_{β_k} in \mathbb{D} for any $k = 0, \dots, m-1$. Now suppose that ℓ_α meets ℓ_{β_k} in \mathbb{D} . It follows that $gh^k = \alpha\beta_k$ is elliptic. If gh^k has infinite order, then S is not semidiscrete and we have case (c). Otherwise we are left with the case where h and gh^k are elliptic elements of finite order. Since $g = gh^k h^{m-k}$, it follows that S is equal to $\langle h, gh^k \rangle$. This means that S is generated by two finite order

elliptic transformations, and so S is a group. This group is equal to the group generated by g and h , and we have case (d).

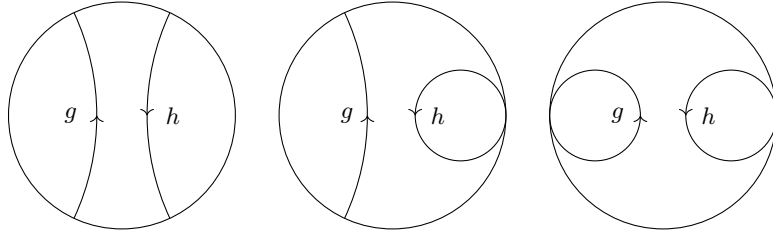
Finally, we suppose $S = \langle g, h \rangle$ is an arbitrary semidiscrete semigroup with nonempty group and inverse free parts. By Lemma 2.2, we know that the group part is generated by just one of the transformations g or h – let us say h . Therefore h must be elliptic of finite order. As $\langle g, h \rangle$ is semidiscrete and has group part $\langle h \rangle$, the element g cannot be elliptic; it must be loxodromic or parabolic. Furthermore, the map gh^k cannot be elliptic for any k , otherwise $\langle h, gh^k \rangle$ is a group and is equal to S . We are now left with the possibility described in case (b). □

We emphasise that in case (d) of the theorem above, S is a group and so S is semidiscrete if and only if the group generated by the elliptic transformations g and gh^k (for appropriate k) is discrete. Whether or not this is the case can be determined by the Gilman-Maskit algorithm detailed in [18].

We now consider two-generator semigroups where both generators are not elliptic and the generators' fixed points do not coincide. It is useful to describe a pair of transformations g, h as *antiparallel* if neither is elliptic, the set of fixed points of g is disjoint from the set of fixed points of h , and no proper closed interval of the unit circle whose end points are fixed points of g or h is mapped strictly within itself by either g or h . Intuitively this means that g and h 'point' in opposite directions. Figure 3.3 shows the three possible configurations of the transformations (up to conjugation in $\text{Aut}(\mathbb{D})$ and replacement of S with S^{-1}). The horoballs in the last two diagrams represent parabolic elements which fix the horoballs setwise, and move points on the horoball in the direction indicated.

The following theorem handles the remaining cases where g and h are not antiparallel.

THEOREM 3.13. *If g and h are both not elliptic, are not antiparallel, and $S = \langle g, h \rangle$ is nonelementary, then S is semidiscrete and inverse free.*

FIGURE 3.3. Possible configurations of antiparallel transformations g and h

PROOF. By first replacing S by S^{-1} if necessary, the generators of S are conjugate to one of the possibilities shown in Figure 3.4.

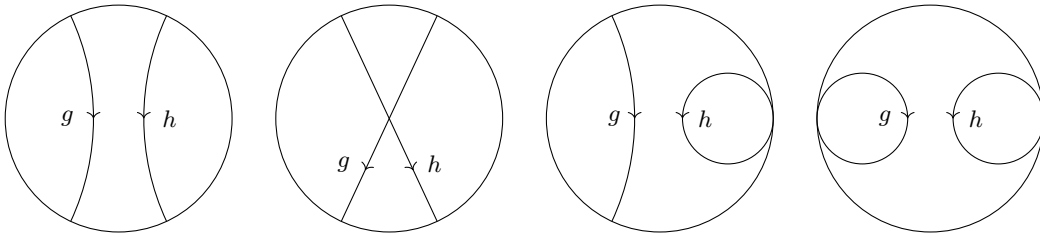


FIGURE 3.4. Possible configurations of generators in Theorem 3.13

In each case it is easy to identify a closed interval contained in the unit circle that is mapped strictly inside itself by both generators, and so S is semidiscrete and inverse free by Theorem 2.5. \square

It remains for us to classify two-generator semigroups that are nonelementary and have antiparallel generators. We give two theorems that can be used as the basis for an algorithm that can determine whether or not a semigroup generated by two antiparallel transformations g and h is semidiscrete, semidiscrete and inverse free, or neither. In principle, given g and h , an upper bound for the runtime of the algorithm can be computed before execution. In fact we show that when g and h are antiparallel, S is semidiscrete and inverse free if and only if S contains no elliptic maps, and in that case the group generated by g and h is itself discrete and elliptic-free. The following lemma records an observation that we shall use repeatedly.

LEMMA 3.14. *Suppose $S = \langle g, h \rangle$ where neither of g, h is an elliptic element of finite order. If S contains an elliptic element of finite order, then S is equal to the group generated by g and h .*

PROOF. Suppose $f \in S$ is elliptic of order $n = 2, 3, \dots$. Written as a word in g and h , f is not a power of g nor a power of h , and so the word contains both the letters g and h . Since $f^n = I$, we have a word in both letters g and h that equals the identity. Without loss of generality we can suppose that f takes the form $f = gf_1$, where f_1 is a nonempty word. Then $(gf_1)^n = I$ and so $g^{-1} = f_1(gf_1)^{n-1} \in S$. Since the letter h features in the word f and $g^{-1} \in S$, we can similarly deduce that $h^{-1} \in S$. It follows that $S = G$. \square

As a consequence of the above, whenever we encounter an elliptic element in S we can check if S is semidiscrete as follows. Either the elliptic element is of infinite order, and so S is not semidiscrete, or the elliptic element has finite order, in which case $S = G$ by Lemma 3.14 and S is semidiscrete exactly when G is discrete. In the latter case we can appeal to the algorithm given in [18] to decide if G is discrete or not.

Our next theorem deals with the more straightforward case where at least one of g or h is parabolic. In order to prove it we shall make use of Poincaré's Polygon Theorem [3, Theorem 9.8.4], which we now describe.

Let P be a *convex polygon*, that is a hyperbolically convex subset of \mathbb{D} whose boundary can be written as the finite union of non-trivial geodesic segments and connected subsets of \mathbb{S}^1 . The geodesic segments may land on the ideal boundary at both ends, one end, or neither end. A *side* of P is a maximal subset of the boundary of P that is either contained within a geodesic segment, or a connected subset of \mathbb{S}^1 . A *vertex* of P is any end point of one of the sides of P , which may lie in \mathbb{S}^1 or in \mathbb{D} . A vertex that lies on \mathbb{S}^1 is called an *ideal vertex*. A *side-pairing transformation* of P is a hyperbolic isometry g_s (not necessarily orientation preserving), that maps one side s of P bijectively to another side t such that $g_s(P) \cap P = t$.

THEOREM 3.15. *Poincaré's Polygon Theorem. Let P be a polygon in \mathbb{D} equipped with a set of side-pairing transformations, g_s , one for each side of P , such that $g_t = g_s^{-1}$ if g_s pairs s with t . Suppose there exists $\epsilon > 0$ with the following properties.*

- (i) *For each vertex $x_0 \in \mathbb{D}$ there are vertices x_1, \dots, x_n of P and hyperbolic isometries f_0, \dots, f_{n+1} such that $f_0 = f_{n+1} = I$ and for each $j = 0, \dots, n$ there is a side s such that $f_{j+1} = f_j g_s$. Further suppose that if $N_j = \{z \in P \mid \rho(x_j, z) < \epsilon\}$ then*

the sets $f_j(N_j)$ do not overlap and have union $B(x_0, \epsilon)$.

- (ii) For each ideal vertex y_0 of P there are ideal vertices y_1, \dots, y_n of P and hyperbolic isometries f_0, \dots, f_{n+1} such that $f_0 = f_{n+1} = I$ and for each $j = 0, \dots, n$ there is a side s of P such that $f_{j+1} = f_j g_s$, and $f_j(N'_j)$ are contiguous and do not overlap, where $N'_j = \{z \in P \mid |y_j - z| < \epsilon\}$.

Then the group G generated by the side-pairing transformations g_s is discrete, P is a fundamental region for G , and each relation of G arises from a vertex in \mathbb{D} and finite composition of its associated maps f_0, \dots, f_{n+1} that fixes P equated with the identity.

Notice that if all the side-pairing transformations for polygon P are orientation preserving, and all the vertices of P are ideal, then the group generated by the side-pairing transformations has no non-trivial relations: it is a free group.

We shall make use of the following lemma in both the subsequent theorem and in Theorem 3.19

LEMMA 3.16. Suppose $g_1 = g$ and $h_1 = h$ are Möbius transformations, and n_j is a sequence of positive integers. Further suppose we have finite sequences g_1, \dots, g_k and h_1, \dots, h_k that satisfy $\{g_{j+1}, h_{j+1}\} = \{g_j^{n_j} h_j, g_j^{n_j+1} h_j\}$ for $j = 1, \dots, k-1$. If the group generated by g_k and h_k is free, then the semigroup $\langle g, h \rangle$ does not contain the identity.

PROOF. Let G_j be the group generated by g_j and h_j . Clearly $G_{j+1} \subseteq G_j$. Since

$$g_j = (g_j^{n_j+1} h_j)(g_j^{n_j} h_j)^{-1}$$

and

$$h_j = g_j^{-n_j} (g_j^{n_j} h_j),$$

we see that $G_j \subseteq G_{j+1}$. Hence $G_j = G_1$ for each j . If a group is free, then it is freely generated by any choice of generators. It follows that since G_k is free, so is G_1 , and in particular, the semigroup generated by g and h does not contain the identity. \square

We remark that the conclusion of the lemma holds if the assumption $\{g_{j+1}, h_{j+1}\} = \{g_j^{n_j} h_j, g_j^{n_j+1} h_j\}$ is replaced with $\{g_{j+1}, h_{j+1}\} = \{h_j g_j^{n_j}, h_j g_j^{n_j+1}\}$.

THEOREM 3.17. *Suppose g and h are antiparallel Möbius transformations, and h is parabolic. Then S is semidiscrete and inverse free if and only if gh^n is not elliptic for any $n \in \mathbb{N}$. Moreover, exactly one of the following is true.*

- (a) *The map gh^n is not elliptic for all $n \in \mathbb{N}$, in which case S is free and the group generated by g and h is discrete and elliptic-free;*
- (b) *for some $n \in \mathbb{N}$ the map gh^n is an elliptic transformation of infinite order and S is not semidiscrete; or*
- (c) *for some $n \in \mathbb{N}$ the map gh^n is an elliptic transformation of finite order, and S is a group that is semidiscrete precisely when S is a discrete group.*

PROOF. We first consider the case where g is parabolic. Let ℓ be the geodesic landing at the fixed points of g and h , and denote the reflection in ℓ by σ . We can choose a geodesic ℓ_g such that $g = \sigma_g \sigma$ where σ_g is the reflection in ℓ_g , and geodesic ℓ_h such that $h = \sigma \sigma_h$ where σ_h is the reflection in ℓ_h . Since g and h are antiparallel, ℓ does not separate ℓ_h from ℓ_g in \mathbb{D} . Since h and g are parabolic, ℓ_g and ℓ_h land at the fixed points of g and h respectively. We have the two cases shown in Figure 3.5: either ℓ_g meets ℓ_h in \mathbb{D} , or not. If ℓ_g and ℓ_h do meet in the unit disc, then $gh = \sigma_g \sigma_h$ fixes the point of intersection, and

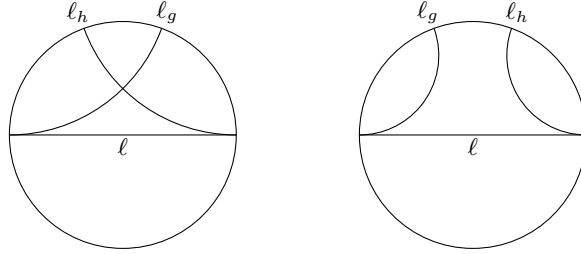


FIGURE 3.5. The two cases of Theorem 3.17 when both g and h are parabolic

so gh is an elliptic map. If gh is elliptic of infinite order, then S is not semidiscrete and we have case (b). Otherwise gh has finite order, in which case Lemma 3.14 tells us that S is a group. Hence S is semidiscrete exactly when S is a discrete group, and we fall into case (c).

Otherwise the geodesics ℓ, ℓ_h and ℓ_g bound a region U in \mathbb{D} , and we show that we have the situation describe by case (a). By Poincaré's Polygon Theorem 3.15, if Γ is the group of conformal and anticonformal Möbius transformations generated by the reflections σ, σ_g

and σ_h , then Γ is discrete and contains the group generated by S . It follows that S is semidiscrete. We note that $U \cup \sigma(U)$ is a fundamental polygon for the group generated by S , which we denote by T , and each vertex lies on the ideal boundary. It follows from Poincaré's theorem that T , and so S itself, does not contain any elliptic elements. Since there are no nontrivial vertex relations, Poincaré's theorem further shows that T is free as a group generated by g and h , and hence S is free as a semigroup generated by g and h . It follows that gh^n is not elliptic for any $n \in \mathbb{N}$. Indeed, S is a Schottky semigroup, and so does not contain the identity. To see this, let J_g be the closed interval on the unit circle whose end points are those of ℓ_g and that does not contain the fixed point of h . Similarly let J_h be the closed interval on the unit circle whose end points are those of ℓ_h and that does not contain the fixed point of g , as shown in Figure 3.6.

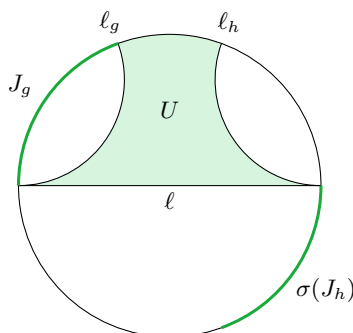


FIGURE 3.6. The domain U with intervals J_g and $\sigma(J_h)$

Both g and h map $J_g \cup \sigma(J_h)$ strictly within itself, and so S is a Schottky semigroup, as claimed.

We now consider the case where g is loxodromic and h is parabolic. Staying in the unit disc model, we conjugate so that i is the fixed point of h and the imaginary axis is perpendicular to the axis of g . We let σ denote the reflection in the imaginary axis, and choose σ_g such that $g = \sigma_g \sigma$. For each $n = 0, 1, 2, \dots$ we let σ_n be the reflection such that $h^n = \sigma \sigma_n$. Note that $\sigma = \sigma_0$. We let ℓ , ℓ_g and ℓ_n denote the lines of reflection of σ , σ_g and σ_n respectively. We discriminate between two cases: either σ_g and σ_n meet in \mathbb{D} for some $n = 0, 1, 2, \dots$, or not. If they do meet, then $gh^n = \sigma_g \sigma_n$ is an elliptic transformation,

and, just as in the case where both g and h were parabolic, we have case (b) if gh^n has infinite order and case (c) if gh^n has finite order.

Otherwise σ_g and σ_n do not meet in \mathbb{D} for each $n = 0, 1, 2, \dots$. It follows that for some n the lines ℓ_n , ℓ_g and ℓ_{n+1} bound a region U , as shown in Figure 3.7. We show that S is described by case (a). By Poincaré's Polygon Theorem 3.15, the group $\Gamma = \langle \sigma_n, \sigma_{n+1}, \sigma_g \rangle$

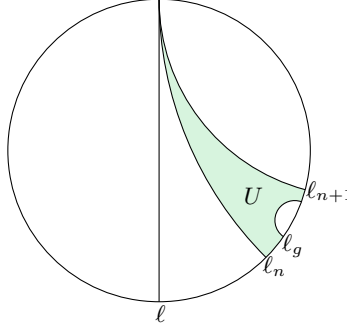


FIGURE 3.7. The lines ℓ , ℓ_n , ℓ_{n+1} and ℓ_g , and the region U

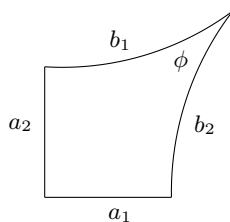
is discrete, with fundamental region U . Moreover since all the vertices of U are ideal, the only relations on Γ are the trivial ones: $\sigma_n^2 = \sigma_{n+1}^2 = \sigma_g^2 = I$. Let T be the group generated by gh^{n+1} and gh^n . Certainly $T \subseteq \Gamma$, and since T is orientation preserving it does not contain σ_n , σ_{n+1} or σ_g . It follows that T has no nontrivial relations, in other words, T is a free group. It now follows from Lemma 3.16 and the remarks following its proof that S does not contain the identity. To see that S is semidiscrete, note that $\sigma_{k-1} = \sigma_k \sigma_{k+1} \sigma_k$ for all $k \in \mathbb{N}$, hence by induction both σ_0 and σ_1 lie in Γ . It follows that both the generators g and h also lie in Γ , and so S is semidiscrete since Γ is discrete. Finally, since the group generated by g and h is orientation preserving and contained in Γ , it must be free, and in particular, contains no elliptic transformations. \square

In the case where h is parabolic and g loxodromic, we remark that it is not necessary to test whether or not gh^n is an elliptic transformations for every $n \in \mathbb{N}$, in order to determine which of the cases (a), (b), or (c) our semigroup $\langle g, h \rangle$ falls into. Indeed, for a particular g and h , the proof gives the basis of an algorithm that identifies the largest such n we must test. If both g and h are parabolic, then $n = 1$ suffices; we only need to

verify whether or not gh is an elliptic map.

Our next task is to treat the case where g and h are both loxodromic and antiparallel. We first require a result on the geometry of quadrilaterals, which can be found in [3, Theorem 7.17.1].

LEMMA 3.18. *Consider the quadrilateral with side lengths b_1 , a_2 , a_1 and b_2 , listed in the anticlockwise sense. Let ϕ denote the angle between b_1 and b_2 , and suppose all the other internal angles are right angles.*



Then we have:

- (i) $\sinh a_1 \sinh a_2 = \cos \phi$, and
- (ii) $\cosh a_1 = \cosh b_1 \sin \phi$.

THEOREM 3.19. *Suppose g, h are two antiparallel loxodromic maps. Then $S = \langle g, h \rangle$ is semidiscrete and inverse free if and only if S is elliptic-free. Moreover, whenever S is semidiscrete and inverse free:*

- (i) *the group generated by g and h is discrete and elliptic-free; and*
- (ii) *we can find two elements in S that generate G , and G is a free group.*

If S does contain an elliptic element, then either S is not semidiscrete, or $S = G$ and G is discrete.

PROOF. We work in the disc model. By applying a conjugation if necessary, we can assume that the axes of g and h are symmetrical about both the real and imaginary axes, as is the case in both examples of Figure 3.8. Let ℓ be the geodesic contained in the imaginary axis that cuts the axes of both g and h orthogonally, and let σ be the reflection in ℓ . There exists a unique geodesic ℓ_g orthogonal to the axis of g such that $g = \sigma_g \sigma$ (where σ_g is the reflection in ℓ_g) and a unique geodesic ℓ_h orthogonal to the axis of h such

that $h = \sigma\sigma_h$ (where σ_h is the reflection in ℓ_h). Since elliptic maps are characterised by the existence of fixed points inside \mathbb{D} , it follows $gh = \sigma_g\sigma_h$ is elliptic exactly when ℓ_g and ℓ_h meet in \mathbb{D} .

For a loxodromic Möbius transformation f , let $\rho(f)$ denote half its translation length and let $\text{ax}(f)$ denote the axis of f . Hence $\rho(g)$ and $\rho(h)$ are the (hyperbolic) distances from ℓ to ℓ_g and from ℓ to ℓ_h , respectively. Without loss of generality we suppose $\rho(h) \geq \rho(g)$. There are two cases to consider, which are shown in Figure 3.8: either ℓ_h meets $\text{ax}(g)$ in \mathbb{D} , or otherwise.

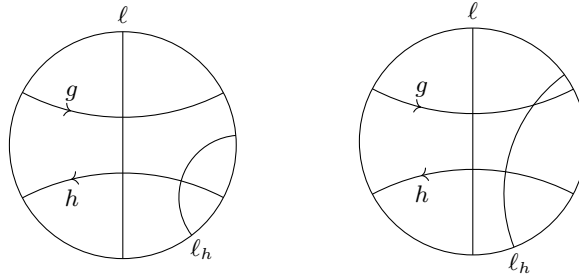


FIGURE 3.8. The cases where $\{g, h\}$ is decidable and otherwise

Suspending for now our assumption that $\rho(h) \geq \rho(g)$, we say that any pair of antiparallel transformations g and h is *decidable* if, after conjugating so that their axes enjoy the symmetry shown in Figure 3.8, either ℓ_h does not meet $\text{ax}(g)$ in \mathbb{D} (they may meet on the ideal boundary) or ℓ_g does not meet $\text{ax}(h)$ in \mathbb{D} . We use the term ‘decidable’ because whenever $\{g, h\}$ has this property, it is relatively easy to determine if S is semidiscrete and inverse free, using a procedure that we now describe. In fact we show that when $\{g, h\}$ is decidable and $\rho(h) \geq \rho(g)$, then S is semidiscrete and inverse free if and only if $g^n h$ is not elliptic for any $n \in \mathbb{N}$. To prove this, we use similar ideas to those used in the proof of Proposition 3.17. For each $n = 0, 1, 2, \dots$ consider the geodesic ℓ_n orthogonal to $\text{ax}(g)$ such that $g^n = \sigma_n \sigma$. Note that $\ell_0 = \ell$, $\ell_1 = \ell_g$ and that the (hyperbolic) distance from ℓ to ℓ_n is $n\rho(g)$. Since $\{g, h\}$ is decidable and $\rho(h) \geq \rho(g)$, then ℓ_h does not meet $\text{ax}(g)$. If for some $n \in \mathbb{N}$ the two geodesics ℓ_n and ℓ_h meet in \mathbb{D} , then $g^n h = \sigma_n \sigma_h$ is elliptic. Otherwise for some least $n = 0, 1, 2, \dots$, the lines ℓ , ℓ_{n+1} and ℓ_h bound a region in \mathbb{D} . We

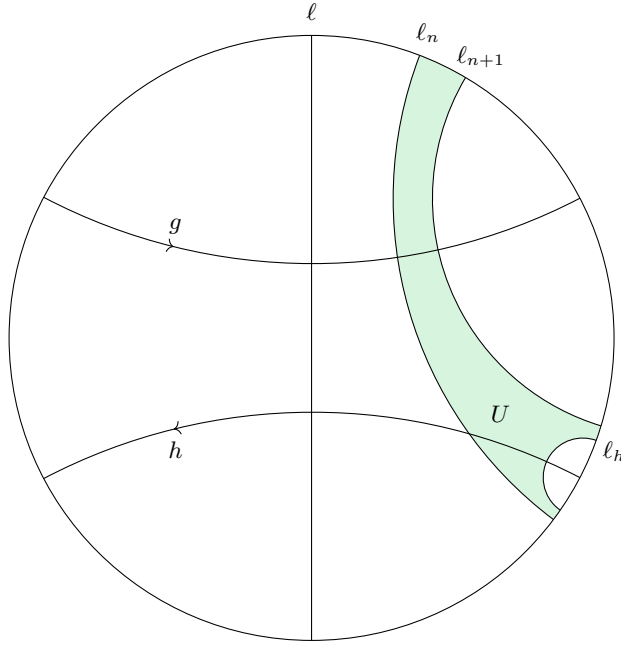


FIGURE 3.9. The lines ℓ , ℓ_n , ℓ_{n+1} and ℓ_h , and the region U , for the decidable case

now apply Poincaré's theorem to the region U bounded by ℓ_n , ℓ_{n+1} and ℓ_h (see Figure 3.9), and infer that the group Γ of conformal and anticonformal maps generated by σ_n , σ_{n+1} and σ_h is discrete. As the orientation-preserving subgroup of Γ , G is discrete and so S is semidiscrete. The region $U \cup \sigma_n(U)$ is a fundamental region for G , and by considering the action of $g^n h$ and $g^{n+1} h$ on $U \cup \sigma_n(U)$, it follows from Poincaré's theorem that G is freely generated as a group by $g^n h$ and $g^{n+1} h$, and moreover contains no elliptic elements. (The semigroup may contain parabolic elements, for example if either of ℓ_n or ℓ_{n+1} meet ℓ_h on the ideal boundary.) It now follows from Lemma 3.16 that $S = \langle g, h \rangle$ does not contain the identity. The preceding discussion can be used as the basis for an algorithm that determines whether or not a decidable pair of antiparallel Möbius transformations generates a semidiscrete and inverse free semigroup. More importantly the above helps to establish a more general algorithm that applies when the transformations are not decidable, which is the case we consider next.

For antiparallel Möbius transformations g and h we define the following quantity

$$d(g, h) = \sinh[\max(\rho(h), \rho(g))] \cdot \sinh[\rho(\text{ax}(g), \text{ax}(h))].$$

By considering the shaded quadrilateral in Figure 3.10, it follows from an application of Lemma 3.18 that ℓ_h does not meet $\text{ax}(g)$ in \mathbb{D} precisely when

$$\sinh[\rho(h)] \cdot \sinh[\rho(\text{ax}(g), \text{ax}(h))] \geq 1.$$

Hence $\{g, h\}$ is decidable precisely when $d(g, h) \geq 1$.

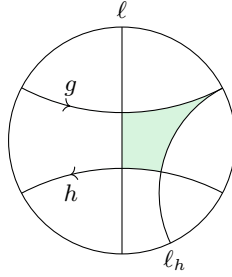


FIGURE 3.10. The case where $d(g, h) = 1$

In the case where g and h are not decidable, we proceed as follows: We let $(g_0, h_0) = (g, h)$ and construct a finite sequence $(g_0, h_0), (g_1, h_1), \dots, (g_m, h_m)$ for some $m \in \mathbb{N}$. For each $k = 0, \dots, m$ the maps g_k and h_k are loxodromic and antiparallel elements of S , and generate G as a group. Suppose (g_k, h_k) has been constructed. We show that either:

- (i) $g_k^n h_k$ is elliptic for some positive integer n ; or
- (ii) at least one of g_k and h_k is parabolic; or
- (iii) $d(g_k, h_k) \geq 1$; or
- (iv) we can find antiparallel $g_{k+1}, h_{k+1} \in S$ that generate G as a group, and satisfy

$$d(g_{k+1}, h_{k+1}) > d(g_k, h_k) / (1 - d(g_k, h_k)^2).$$

If at any stage in the construction we encounter case (i), then S is not semidiscrete and inverse free. In case (ii) we can appeal to Proposition 3.17 and conclude that either S contains an elliptic map, or the group generated by g_k and h_k , which by construction is G itself, is discrete and freely generated by g_k and h_k . In the latter case, by invoking

Lemma 3.16, we see that $I \notin S$ and therefore S is semidiscrete and inverse free. Similarly in case (iii) we have $d(g_k, h_k) \geq 1$, so that $\{g, h\}$ is decidable. This means we can use the procedure described above to either conclude that S contains an elliptic element, or that G is discrete and freely generated by g_k and h_k . Hence S is semidiscrete and, by Lemma 3.16, is also inverse free. In case (iv) we can continue the sequence and find antiparallel loxodromic maps $g_{k+1}, h_{k+1} \in S$ that generate G as a group. The sequence must terminate because if d_k is a sequence such that $0 < d_k < 1$ and $d_{k+1} > d_k/(1 - d_k^2)$, then d_k cannot remain in $(0, 1)$ for all k . Moreover if m is the least positive integer such that d_m is not in $(0, 1)$, then $d_m \geq 1$. This means that we cannot continue constructing the sequence (g_k, h_k) indefinitely without meeting one of the cases (i), (ii) or (iii).

If at any point of the algorithm we encounter an elliptic element, it is of finite or infinite order. If the elliptic element has infinite order then S is not semidiscrete. Otherwise the elliptic element has finite order, and so $S = G$ by Lemma 3.14; and so S is semidiscrete if and only if G is discrete. A finite time algorithm that determines whether or not G is discrete is given in [18]. Conversely if G is discrete then every elliptic element in G has finite order, and hence the same is true of S . This proves the last sentence in the statement of the theorem.

It remains to show that (g_k, h_k) can be constructed as claimed. For clarity of exposition we take $k = 0$ so that $g = g_0$ and $h = h_0$. As in the decidable case, we let ℓ_n be the geodesic orthogonal to $\text{ax}(g)$ such that $g^n = \sigma\sigma_n$, and so $g^n h = \sigma_h\sigma_n$. If ℓ_n meets ℓ_h for some $n \in \mathbb{N}$, then $g^n h$ is elliptic and we have case (i). Another possibility is that ℓ_h meets either of ℓ_n or ℓ_{n+1} on the ideal boundary. In this case $g^n h$ and $g^{n+1} h$ can be seen to be antiparallel, and are either both parabolic, or one is parabolic and one loxodromic. In these cases we can appeal to Proposition 3.17 and infer that either S contains an elliptic map, or G is discrete and elliptic-free. Otherwise for some n , ℓ_h separates ℓ_n from ℓ_{n+1} , and $g^n h$ and $g^{n+1} h$ are both loxodromic (see Figure 3.11). We let $(g_1, h_1) = (g^n h, g^{n+1} h)$ and set n_1 equal to n . Since $g^n h = \sigma_n \sigma_h$, its axis is orthogonal to both ℓ_n and ℓ_h . Similarly $g^{n+1} h = \sigma_{n+1} \sigma_h$ so that its axis is orthogonal to ℓ_{n+1} and ℓ_h . It can be seen from

the geometry that g^nh and $g^{n+1}h$ are antiparallel. To verify this, observe that $\text{ax}(g^nh)$ is the mutual perpendicular of ℓ_n and ℓ_h , and, since $g^nh = \sigma_n\sigma_h$, we have g^nh parallel with h . Similarly $g^{n+1}h$ is parallel with g . Since the axes of g^nh and $g^{n+1}h$ are both orthogonal to ℓ_h , if they were to meet in $\overline{\mathbb{D}}$ then they must be equal setwise, so we must have $\sigma_h(\ell_n) = \ell_{n+1}$. But this cannot be, for then $\text{ax}(g)$, as the mutual perpendicular of ℓ_n and ℓ_{n+1} , must be perpendicular to ℓ_h . (This is more clear upon considering Figure 3.11 and conjugating so that ℓ_h is a Euclidean diameter.) Since $g = (hg^n)^{-1}(hg^{n+1})$, the element g belongs to the group generated by g^nh and $g^{n+1}h$, and so h also belongs to this group. Hence g^nh and $g^{n+1}h$ generate G as a group.

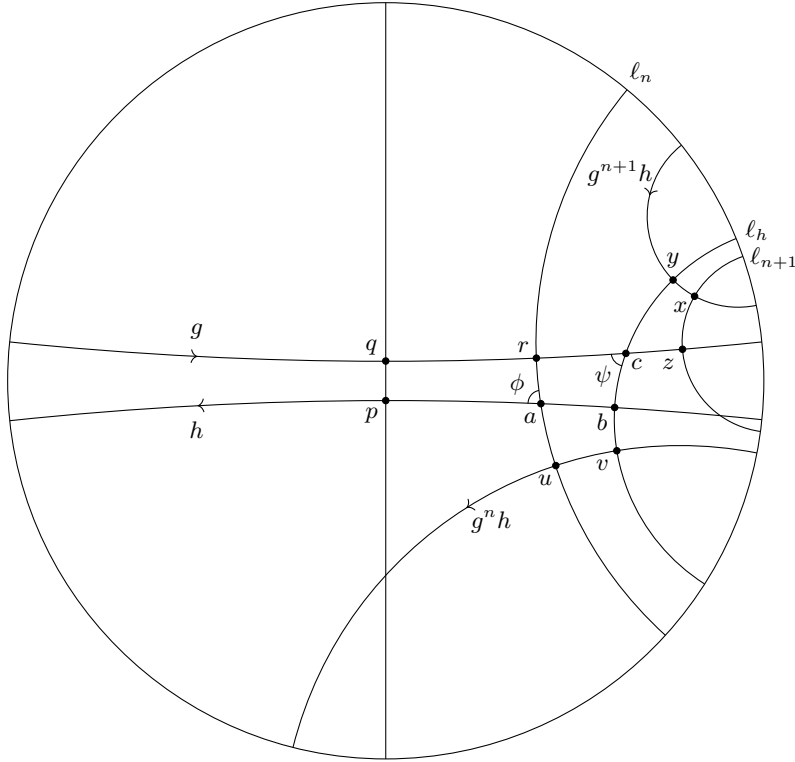


FIGURE 3.11. The geometry of constructing (g_1, h_1) from (g, h)

For points α and β in \mathbb{D} we let $[\alpha, \beta]$ denote the geodesic line joining α to β , and we abbreviate $\rho(\alpha, \beta)$ to $\rho_{\alpha\beta}$. Notice that $[u, v]$ is contained in $\text{ax}(g^{n+1}h)$ and $[y, x]$ is contained

in $\text{ax}(g^n h)$. With $g_1 = g^n h$ and $h_1 = g^{n+1} h$ we want to find

$$d(g_1, h_1) = \sinh[\max(\rho(g^{n+1} h), \rho(g^n h))] \cdot \sinh[\rho(\text{ax}(g^n h), \text{ax}(g^{n+1} h))].$$

We have $\rho(\text{ax}(g^n h), \text{ax}(g^{n+1} h)) = \rho_{cy} + \rho_{cb} + \rho_{bv}$ and $\rho(g^{n+1} h) = \rho_{xy} \leq \max(\rho(g^{n+1} h), \rho(g^n h))$.

Hence to prove $d(g_1, h_1) > d(g_0, h_0)/(1 - d(g_0, h_0)^2)$ it suffices to show that

$$(5) \quad \sinh[\rho_{cy} + \rho_{cb} + \rho_{bv}] \sinh[\rho_{xy}] > d_0/(1 - d_0^2).$$

In order to proceed we consider the geometry of several hyperbolic quadrilaterals found within the figure above and repeatedly apply Lemma 3.18. Considering the quadrilateral $czyx$ we have

$$(6) \quad \cosh \rho_{cy} \sin \psi = \cosh \rho_{zx}$$

and

$$(7) \quad \sinh \rho_{zx} \sinh \rho_{xy} = \cos \psi.$$

Considering the quadrilateral $pqcb$ we have

$$(8) \quad \cosh \rho = \cosh \rho_{cb} \sin \psi,$$

where $\rho = \rho(\text{ax}(g), \text{ax}(h))$. Finally, considering the quadrilateral $abuv$ we obtain

$$(9) \quad \cos \phi = \sinh \rho_{uv} \sinh \rho_{bv}.$$

Equations (6) and (7) yield

$$(10) \quad \cosh \rho_{cy} = \frac{1}{\sin \psi} \sqrt{\frac{\cos^2 \psi}{\sinh^2 \rho_{xy}} + 1}$$

and so

$$(11) \quad \sinh \rho_{cy} = \frac{1}{\tan \psi \tanh \rho_{xy}}.$$

We can now express $\sinh(\rho_{cy} + \rho_{cb} + \rho_{bv})$ in terms of \sinh or \cosh of ρ_{cy}, ρ_{cb} and ρ_{bv} , and eliminate these terms using equations (11), (8) and (9) above.

We have

$$\begin{aligned}
\sinh(\rho_{cy} + \rho_{cb} + \rho_{bv}) &= \sinh \rho_{cy} (\cosh \rho_{cb} \cosh \rho_{bv} + \sinh \rho_{cb} \sinh \rho_{bv}) + \\
&\quad \cosh \rho_{cy} (\sinh \rho_{cb} \cosh \rho_{bv} + \sinh \rho_{bv} \cosh \rho_{cb}) \\
&\geq \sinh \rho_{cy} \cosh \rho_{cb} \cosh \rho_{bv} \\
&= \frac{1}{\tan \psi \tanh \rho_{xy}} \frac{\cosh \rho}{\sin \psi} \sqrt{\frac{\cos^2 \phi}{\sinh^2 \rho_{uv}} + 1} \\
&\geq \frac{\cosh \rho_{xy} \cosh \rho \cos \psi}{\sin^2 \psi \sinh \rho_{xy}},
\end{aligned}$$

and so

$$\begin{aligned}
\sinh(\rho_{cy} + \rho_{cb} + \rho_{bv}) \sinh \rho_{xy} &> \frac{d_0}{(1 - d_0^2)} \cosh \rho_{xy} \cosh \rho \\
&> \frac{d_0}{(1 - d_0^2)}
\end{aligned}$$

as required. \square

It follows from Theorem 3.19 that if g and h are antiparallel loxodromic Möbius transformations, then $S = \langle g, h \rangle$ is semidiscrete if and only if S is contained in a discrete group. This is not true when g and h are not antiparallel. Some of the ideas used in the proof above are taken from [18]; however, the proof given there uses the fact that if the group generated by g and h is discrete and nonelementary, then Jørgensen's inequality (see [3, Theorem 5.4.1]) is satisfied. That is,

$$|\mathrm{tr}^2(g) - 4| + |\mathrm{tr}(ghg^{-1}h^{-1}) - 2| \geq 1.$$

We could not assume Jørgensen's inequality, since it was not clear beforehand that if S is semidiscrete and generated by two antiparallel loxodromic Möbius transformations, then S is contained in a discrete group. Neither was it clear beforehand how to determine whether or not S contains an elliptic map; indeed, if S does contain an elliptic map, then the proof of Theorem 3.19 can be used to find one such map explicitly.

Theorem 3.19 also has relevance to the study of groups, for it serves as a test to determine whether or not the group generated by two loxodromic transformations with disjoint axes

is discrete and elliptic-free, without making use of Jørgensen's inequality.

We now collate results obtained in this section into the following theorem.

THEOREM 3.20. *Suppose the semigroup $S = \langle g, h \rangle$ is nonelementary. Then S is semidiscrete if and only if:*

- (i) *exactly one of g and h is elliptic (h say), h has finite order, and gh^k is not elliptic for each $k \in \{0, \dots, \text{order}(h) - 1\}$; or*
- (ii) *S is a Fuchsian group; or*
- (iii) *g and h are not elliptic and not antiparallel; or*
- (iv) *g and h are antiparallel, and S is elliptic-free.*

These cases are mutually exclusive. The results in this section and their proofs collectively constitute a basis for an algorithm that decides whether or not g and h generate a semidiscrete semigroup S . Moreover, it can be determined into which of these four cases S belongs. Theorem 3.20 also classifies those nonelementary two-generator semigroups that are semidiscrete *and inverse free* – these are exactly the cases (iii) and (iv). Case (i) describes a rather small class of semigroups, and is dealt with in Theorem 3.12.

PROOF. Since S is nonelementary the set of fixed points of g is disjoint from the set of fixed points of h . Suppose now that at least one of g and h is elliptic. If both g and h are elliptic and one has infinite order, then S is not semidiscrete. Otherwise both have finite order and $S = G$. Hence S is semidiscrete precisely when G is discrete, which can be determined by the Gilman-Maskit algorithm given in [18]. If exactly one generator is elliptic then we examine cases (a)–(d) in Theorem 3.12. In cases (a) and (c), S is not semidiscrete. Case (b) exactly describes case (i) above. In case (d), both g and h are elliptic elements of finite order and S is a group. Either S is a Fuchsian group, which is case (ii), or otherwise S is not discrete, and hence S is not semidiscrete.

It remains to consider the cases where both generators are either parabolic or loxodromic and have disjoint sets of fixed points. We consider separately the cases where g and h

are antiparallel, and otherwise. If g and h are not antiparallel then we have the situation described in Theorem 3.13. Otherwise g and h are antiparallel, and the procedure described in the proof of Theorem 3.17 (when at least one of g or h is parabolic), or Theorem 3.19 (when g and h are both loxodromic), can decide whether or not S is semidiscrete, semidiscrete and inverse free, or neither. If S is not semidiscrete and inverse free then the procedure identifies an elliptic transformation in S . If this elliptic transformation is of infinite order then S is not semidiscrete. Otherwise $S = G$ and we must again defer to the Gilman-Maskit algorithm detailed in [18] in order to determine whether or not G is discrete. \square

4. Proof of Theorem 3.2

In this section we prove Theorem 3.2 as well as other results that we shall need later. Our first theorem has a well known counterpart in the theory of Fuchsian groups. Recall that α_f and β_f denote the attracting and repelling fixed points (respectively) of a loxodromic Möbius transformation f .

THEOREM 3.21. *Suppose that S is a semigroup of Möbius transformations that includes loxodromic elements and is not elementary of infinite type. Then for every pair of open subsets U and V of \mathbb{S}^2 such that U meets $\Lambda^+(S)$ and V meets $\Lambda^-(S)$, there is a loxodromic element of S with attracting fixed point in U and repelling fixed point in V .*

PROOF. First suppose that S is elementary. Then by Proposition 1.7, S is one of the types (c), (d), (e) or (f), as stated in the proposition. By inspection and using Theorem 1.8 (i), it can be seen that each of these elementary semigroups satisfies the conclusion of this theorem.

Now suppose that S is not elementary. By Theorem 1.8, there are loxodromic maps f and g in S such that $\alpha_f \in U$ and $\beta_g \in V$, and since $\Lambda^+(S)$ is perfect we can assume that $\alpha_f \neq \beta_g$. If either $\alpha_g \in U$ or $\beta_f \in V$, then we have found a loxodromic map of the required type. Suppose instead that $\alpha_g \notin U$ and $\beta_f \notin V$. For the moment, let us assume also that $\beta_f \neq \alpha_g$ (so that no two of $\alpha_f, \beta_f, \alpha_g$ and β_g are equal). We can choose pairwise disjoint open intervals A_f, B_f, A_g and B_g such that $\alpha_f \in A_f, \beta_f \in B_f, \alpha_g \in A_g, \beta_g \in B_g$, and moreover $A_f \subsetneq U$ and $B_g \subsetneq V$. Now let n be a sufficiently large positive integer

such that f^n maps the complement of B_f into A_f , and f^{-n} maps the complement of A_f into B_f . Suppose also that n is large enough such that g^n maps the complement of B_g into A_g , and g^{-n} maps the complement of A_g into B_g . One can now check that the map $h = f^n g^n$ satisfies $h(A_f) \subsetneq A_f$ and $h^{-1}(B_g) \subsetneq B_g$. Hence h is a loxodromic element of S that satisfies $\alpha_h \in U$ and $\beta_h \in V$, as required.

It remains to consider the case where $\beta_f = \alpha_g$ (and, as before, α_f, β_f and β_g are pairwise distinct). Since S is nonelementary there is an element s of S that maps β_f to a point outside the set $\{\alpha_f, \beta_f, \beta_g\}$. Let $k_n = sg^n$. Then k_n is loxodromic for sufficiently large values of n . Moreover, $\alpha_{k_n} \rightarrow s(\alpha_g)$ and $\beta_{k_n} \rightarrow \beta_g$ as $n \rightarrow \infty$. Choose n such that $\alpha_f, \beta_f, \alpha_{k_n}$ and β_{k_n} are pairwise distinct, and $\beta_{k_n} \in V$. We can now apply the argument of the preceding paragraph with k_n replacing g to obtain a loxodromic element of S with the desired properties. \square

The assumption that S includes loxodromic elements and is not elementary of infinite type, is equivalent to assuming that S is not an elementary semigroup of type (a), (b) or (g), as described in Proposition 1.7. In particular Theorem 3.21 holds for all nonelementary semigroups.

A *Gromov hyperbolic metric space* is a metric space within which we can draw triangles and these triangles are always ‘thin’. More precisely, a Gromov hyperbolic metric space enjoys the following properties. Firstly, every two points in the space are the end points of some minimizing geodesic. This means that for each set of three distinct points in the space we can associate a ‘triangle’ whose vertices are the three points, and whose three edges are the minimizing geodesics between each pair of vertices. Finally, there exists $\delta > 0$ such that for each triangle every point on one of its sides lies a distance at most δ from some point on one of the other two sides. In [12, Proposition 7.4.7], the authors show that an analogous version of Theorem 3.21 holds when S is a semigroup of isometries of a Gromov hyperbolic metric space. The same paper also gives a generalisation of Theorem 1.4 (see [12, Proposition 7.3.1]).

Let us now consider three special families of Möbius transformations in $\text{Aut}(\mathbb{H}^2)$, namely

- (i) $z \mapsto (az - b)/(bz + a)$, $a^2 + b^2 = 1$;
- (ii) $z \mapsto \lambda z$, $\lambda \geq 1$;
- (iii) $z \mapsto z + \mu$, $\mu \geq 0$.

Each of these is a one-parameter semigroup. The first consists of all elliptic rotations about i , the second consists of loxodromic transformations with attracting fixed point ∞ and repelling fixed point 0 , and the third is a collection of parabolic transformations that fix ∞ (and the identity is contained in each family too). We will prove that, up to conjugacy, any closed semigroup that is not semidiscrete contains one of these families. However, there is a caveat here: we must allow conjugacy not just by elements of $\text{Aut}(\mathbb{H}^2)$, but by Möbius transformations that transpose the upper and lower half-planes. Elements of $\text{Aut}(\mathbb{H}^2)$ are maps of the form $z \mapsto (az + b)/(cz + d)$, where $a, b, c, d \in \mathbb{R}$ and $ad - bc > 0$; we also allow conjugation by maps of the same form but with negative determinant, that is $ad - bc < 0$. We shall denote the collection of these maps by $\text{Aut}(\overline{\mathbb{R}})$, since each such transformation fixes the extended real line. The reason we need to conjugate by transformations in $\text{Aut}(\overline{\mathbb{R}})$, rather than just $\text{Aut}(\mathbb{H}^2)$, is to simplify the treatment of case (iii); after all, the two semigroups $\{z \mapsto z + \mu \mid \mu \geq 0\}$ and $\{z \mapsto z + \mu \mid \mu \leq 0\}$ are conjugate by $z \mapsto -z$, but they are not conjugate in $\text{Aut}(\mathbb{H}^2)$. In the next lemma, and indeed throughout this thesis, we write \overline{S} to mean the closure of S in $\text{Aut}(\overline{\mathbb{R}^3})$.

LEMMA 3.22. *Let S be a closed semigroup that is not semidiscrete. Then S is conjugate by an element in $\text{Aut}(\overline{\mathbb{R}})$ to a semigroup that contains one of the families (i), (ii) or (iii).*

PROOF. As S is not semidiscrete, there is a sequence g_n in $S \setminus \{I\}$ that converges uniformly to I , the identity. Suppose first that this sequence contains infinitely many loxodromic elements. By passing to a subsequence, we can assume that every map g_n is loxodromic. Let α_n and β_n be the attracting and repelling fixed points of g_n , respectively. By passing to a further subsequence of g_n , we can assume that the sequences α_n and β_n both converge. Suppose for now that they converge to distinct values, which, after conjugating S , we can assume are ∞ and 0 , respectively. Let h_n be any sequence of Möbius transformations that satisfies $h_n(0) = \alpha_n$, $h_n(\infty) = \beta_n$ and $h_n \rightarrow I$ as $n \rightarrow \infty$. Define $k_n = h_n^{-1}g_nh_n$; this map has repelling fixed point 0 and attracting fixed point ∞ ,

so $k_n(z) = \lambda_n z$, where $\lambda_n > 1$. Furthermore, $k_n \rightarrow I$ as $n \rightarrow \infty$ (so $\lambda_n \rightarrow 1$). Now select any number $\lambda > 1$, and define $f(z) = \lambda z$. By passing to yet another subsequence of g_n if necessary, we can assume that $\lambda_n < \lambda$ for $n = 1, 2, \dots$. For each positive integer n , the sequence $\lambda_n, \lambda_n^2, \lambda_n^3, \dots$ is strictly increasing with limit ∞ . Define t_n to be the unique positive integer such that $\lambda \in [\lambda_n^{t_n}, \lambda_n^{t_n+1})$. Let $f_n = k_n^{t_n}$. Notice that $f_n(1) = \lambda_n^{t_n}$ and $f(1) = \lambda$. Also,

$$\chi(f_n(1), f(1)) \leq \chi(k_n^{t_n}(1), k_n^{t_n+1}(1)) \leq \sigma(k_n^{t_n}, k_n^{t_n+1}) = \sigma(I, k_n).$$

Since f_n fixes 0 and ∞ for each n , we see that $f_n(1) \rightarrow f(1)$, $f_n(0) \rightarrow f(0)$ and $f_n(\infty) \rightarrow f(\infty)$ as $n \rightarrow \infty$, so $f_n \rightarrow f$. Therefore, in this case, the family of type (ii) is contained in our semigroup.

Near the start of the preceding argument we assumed that the sequences α_n and β_n converged to *distinct* values. Let us resume the argument from that point, but this time assume that α_n and β_n converge to the same value, which, after conjugating S if need be, we can assume is ∞ . Let h_n be any sequence of Möbius transformations that satisfies $h_n(\infty) = \beta_n$ and $h_n \rightarrow I$ as $n \rightarrow \infty$. Define $k_n = h_n^{-1}g_nh_n$; this map has attracting fixed point ∞ . Let δ_n be the repelling fixed point of k_n . Then $\delta_n \rightarrow \infty$ as $n \rightarrow \infty$. By passing to a subsequence of g_n if need be, we can assume that the numbers δ_n all have the same sign, which, after conjugating by $z \mapsto -z$ if need be, we can assume is negative. We can write $k_n(z) = \lambda_n(z - \delta_n) + \delta_n$, where $\lambda_n > 1$. As before, $k_n \rightarrow I$ as $n \rightarrow \infty$ (so $\lambda_n \rightarrow 1$). Now select any number $\mu > 0$, and define $f(z) = z + \mu$. By passing to yet another subsequence of g_n if necessary, we can assume that $k_n(0) = \delta_n(1 - \lambda_n) < \mu$ for $n = 1, 2, \dots$. For each positive integer n , the sequence $k_n(0), k_n^2(0), k_n^3(0), \dots$ is strictly increasing with limit ∞ . Define t_n to be the unique positive integer such that $\mu \in [k_n^{t_n}(0), k_n^{t_n+1}(0))$. Let $f_n = k_n^{t_n}$. Then

$$\chi(f_n(0), f(0)) \leq \chi(k_n^{t_n}(0), k_n^{t_n+1}(0)) \leq \sigma(I, k_n).$$

Hence $f_n(0) \rightarrow f(0)$; that is, $\delta_n(1 - \lambda_n^{t_n}) \rightarrow \mu$. Since $\delta_n \rightarrow \infty$, we see that $\lambda_n^{t_n} \rightarrow 1$. Hence, for any real number x ,

$$f_n(x) = \lambda_n^{t_n}x + \delta_n(1 - \lambda_n^{t_n}) \rightarrow x + \mu \quad \text{as } n \rightarrow \infty.$$

So $f_n \rightarrow f$. Therefore, in this case, the family of type (iii) is contained in our semigroup. We began this proof by choosing a sequence g_n in $S \setminus \{I\}$ such that $g_n \rightarrow I$, and supposing that there were infinitely many loxodromic maps in this sequence. If there are infinitely many parabolic maps in the sequence, then we can carry out an argument similar to those we have given already to show that a conjugate of S contains the family of type (iii). The remaining possibility is that almost all the maps g_n are elliptic, in which case we may as well assume that they are all elliptic. If any one of them is elliptic of infinite order, then we obtain the family of type (i) (up to conjugacy) by considering the closure of the semigroup generated by that map alone. If infinitely many of the maps g_n share a fixed point, then it is straightforward to see that \bar{S} is conjugate to a semigroup that contains the family of type (i). The only other possibility is that the maps g_n have infinitely many different fixed points. By passing to a subsequence of g_n , we can assume that the fixed points of g_n are pairwise distinct. Let $h_n = g_n g_{n+1} g_n^{-1} g_{n+1}^{-1}$. Since the maps g_n are of finite order, h_n is an element of S , and $h_n \rightarrow I$ as $n \rightarrow \infty$. Furthermore, one can easily check that h_n is loxodromic, so by the earlier arguments we see that the closure of our semigroup contains one of the families of types (ii) or (iii). \square

For an elliptic map $h \in \text{Aut}(\mathbb{H}^2)$ let $\text{fix}(h)$ denote its unique fixed point in \mathbb{H}^2 and let $\theta(h) \in [0, 2\pi)$ be its angle of rotation. Each of the families (i), (ii) and (iii) can be parametrised by one real parameter. In the case of (i), we use the angle of rotation of the elliptic map, while in (ii) and (iii) we use λ and μ respectively. In order to simplify notation we let t denote the parameter in each family and let f_t denote element in the family, where it is understood that $t \in \mathbb{R}$ in family (i), $t \geq 1$ in (ii) and $t \geq 0$ in (iii).

LEMMA 3.23. *Suppose that g is a loxodromic Möbius transformation with $0 < \alpha_g < \beta_g < +\infty$. Then in each of the families (i), (ii) and (iii) there is an interval $(a, b) \subset \mathbb{R}$ such that $f_t g$ is elliptic for all $t \in (a, b)$. Moreover the functions $t \mapsto \text{fix}(f_t g)$ and $t \mapsto \theta(f_t g)$ are continuous on (a, b) , and the former is injective.*

PROOF. There is an algebraic proof of this lemma, but the geometric proof we offer is more illuminating. We consider cases (i), (ii) and (iii) in turn; first case (i). Let γ denote the reflection in the hyperbolic line ℓ_γ that passes through i and is orthogonal to

the axis of g . Then $g = \gamma\beta$, where β is the reflection in another line ℓ_β that is parallel to ℓ_γ . Let ℓ_t denote the line that passes through i such that $f_t = \alpha_t\gamma$, where α_t is the reflection in ℓ_t . For some value of t , $t = \tau$ say ($\tau = \pi$ suffices), ℓ_τ intersects ℓ_β at a single point in \mathbb{H}^2 . Moreover, $f_\tau g = \alpha_\tau\beta$ is elliptic, because ℓ_τ and ℓ_β intersect at a point in \mathbb{H}^2 . Since the map $t \mapsto f_t$ is continuous, the Möbius group is a topological group, and the set of elliptic transformations is an open subset of the Möbius group, it follows that there is some open interval $(a, b) \subset \mathbb{R}$ containing τ such that $f_t g$ is elliptic for all $t \in (a, b)$. Moreover, since the functions $h \mapsto \text{fix}(h)$ and $h \mapsto \theta(h)$ are continuous, it follows that the maps $t \mapsto \text{fix}(f_t g)$ and $t \mapsto \theta(f_t g)$ are also continuous. Indeed, the continuity is geometrically clear since $\text{fix}(f_t g)$ is the point where ℓ_β and ℓ_t intersect and $\theta(f_t g)$ is twice the anticlockwise angle between ℓ_β and ℓ_t . It is also clear from the geometry that both functions are injective, and we verify this for the function $t \mapsto \text{fix}(f_t g)$. Suppose, then, that a point $\zeta \in \mathbb{H}^2$ is the common fixed point of $f_t g$ and $f_s g$, for some $t, s \in (a, b)$. This means that f_{t-s} fixes $g(\zeta)$. If $t \neq s$ then we must have $g(\zeta) = i$, which is not possible since $\zeta \in \ell_\beta$, $i \in \ell_\gamma$ and $g(\ell_\beta)$ does not meet ℓ_γ . Hence $g(\zeta) \neq i$, and so $t = s$ after all.

Now for case (ii). This time let ℓ_γ be the line that is orthogonal to the axis of g and to the vertical hyperbolic line L between 0 and ∞ (the positive imaginary axis). In Euclidean terms, ℓ_γ is a Euclidean semicircle centred on the origin that is symmetric about L . Outside this semicircle is contained the hyperbolic line ℓ_β that is orthogonal to the axis of g and satisfies $g = \gamma\beta$ (where β and γ are the reflections in ℓ_β and ℓ_γ respectively). Let α_t denote the reflection in the line ℓ_t , chosen such that ℓ_t is orthogonal to L and such that $f_t = \alpha_t\gamma$. Note that f_t , which belongs to family (ii), has attracting fixed point ∞ and repelling fixed point 0, and so for $t > 0$ the line ℓ_t is outside the semicircle bounded by the real axis and ℓ_γ . Note that ℓ_β is not orthogonal to L , since ℓ_γ is the unique geodesic that is orthogonal to both L and the axis of g . It follows that we can choose a value of t , $t = \tau$ say, such that ℓ_t meets ℓ_β at exactly one point in \mathbb{H}^2 . Hence at $t = \tau$, the transformation $f_t g = \alpha_t\beta$ is elliptic and fixes the point in the hyperbolic plane where ℓ_t cuts ℓ_β . As in the case of family (i), we can choose an open interval $(a, b) \subset \mathbb{R}^+$ containing τ such that $f_t g$ is an elliptic transformation for all $t \in (a, b)$. Again, since the functions

$h \mapsto \text{fix}(h)$ and $h \mapsto \theta(h)$ are continuous, the functions $t \mapsto \text{fix}(f_t g)$ and $t \mapsto \theta(f_t g)$ are also continuous on (a, b) . The former function is injective, for if $a < s \leq t < b$ and $f_s g$ and $f_t g$ both fix the same point $\zeta \in \mathbb{H}^2$, then $f_s^{-1} f_t = f_{t-s}$ fixes $g(\zeta)$. If $t \neq s$ then f_{t-s} is loxodromic and so cannot fix any point in \mathbb{H}^2 , hence we must have $t = s$, as required.

Case (iii) is similar to case (ii), and we omit the argument. \square

We now set about proving Theorem 3.2. We will need the following result of Bárány, Beardon and Carne [2, Theorem 3], mentioned in the introduction.

THEOREM 3.24. *Suppose that f and g are noncommuting elements of $\text{Aut}(\mathbb{D})$, and one of them is an elliptic element of infinite order. Then $\langle f, g \rangle$ is dense in $\text{Aut}(\mathbb{D})$.*

We also need the following lemma, which states two separate results. The first follows from the characterisation of elementary groups given on page 84 of [3], while the second follows from [3, Theorem 8.4.1].

LEMMA 3.25.

- (i) *If an elementary group contained in $\text{Aut}(\mathbb{D})$ contains elliptic transformations, then they all fix a common point in \mathbb{D} .*
- (ii) *If a nonelementary subgroup of $\text{Aut}(\mathbb{D})$ contains elliptic transformations whose fixed points accumulate in \mathbb{D} , then the group contains an elliptic transformation of infinite order.*

The next theorem is a stronger version of Theorem 3.2 (we will need this stronger statement later). Recall from the introduction that $\text{Möb}(J)$ denotes the semigroup of those Möbius transformations that map a closed interval J within itself.

THEOREM 3.26. *Suppose that S is a semigroup that is not elementary, semidiscrete, or contained in $\text{Möb}(J)$, for any nontrivial closed interval $J \subset \mathbb{S}^1$. Then there is a two-generator semigroup within S that is dense in $\text{Aut}(\mathbb{D})$.*

PROOF. We frame the proof using the upper half-plane model. By Lemma 3.22, we can assume, after conjugating S by a Möbius transformation in $\text{Aut}(\overline{\mathbb{R}})$, that its closure

\overline{S} contains one of the families of maps (i), (ii) or (iii). In order to apply Lemma 3.23, we will prove that there is a loxodromic map g in S with $0 < \alpha_g < \beta_g < +\infty$ (possibly after conjugating S again). To this end, suppose first that \overline{S} contains the family of type (i). The sets $\Lambda^-(S)$ and $\Lambda^+(S)$ are perfect, and, since S is not elementary, by Theorem 3.21 we can choose a loxodromic map g in S such that $\chi(\alpha_g, \beta_g)$ is less than $\chi(0, \infty) = 2$, the maximum value of χ . After conjugating S by elliptic rotations about i , and possibly the map $z \mapsto -z$ (which fixes the family of type (i)), we can assume that $0 < \alpha_g < \beta_g < +\infty$. Suppose now that \overline{S} contains a family of type (ii); then $0 \in \Lambda^-(S)$ and $\infty \in \Lambda^+(S)$. One possibility might be that $\Lambda^-(S)$ is contained in a closed interval J , and $\Lambda^+(S)$ is contained in the closed interval $\overline{\mathbb{R}} \setminus J$. In fact this cannot be: for then there exists a unique complementary connected component of Λ^- whose interior meets Λ^+ . Let us denote this component, regarded as a subset of $\overline{\mathbb{R}}$, by U . Since S is not elementary, U has a non-empty interior and $\overline{U} \neq \overline{\mathbb{R}}$. Since elements of S map complementary components of Λ^- into complementary components of Λ^- , we see that U is forward invariant under S , and so $S \subset \text{Möb}(\overline{U})$ contrary to our assumption. It follows, then, that there exist points p in $\Lambda^-(S)$ and q in $\Lambda^+(S)$ with either $-\infty < p < q < 0$ or $0 < q < p < +\infty$. After conjugating S by the map $z \mapsto -z$ (which fixes the family of type (ii)), we can assume that the points p and q satisfy $0 < q < p < +\infty$. We can now apply Theorem 3.21 to deduce the existence of a loxodromic map g in S with $0 < \alpha_g < \beta_g < +\infty$.

The remaining case is that \overline{S} contains a family of type (iii). This can be dealt with in a similar way to the previous case, only this time if we have $-\infty < p < q < 0$ and not $0 < q < p < +\infty$, we must conjugate by a parabolic map of the form $z \mapsto z + t$ (for some $t \in \mathbb{R}$) that fixes the family (iii). Once again we deduce the existence of a loxodromic map g with $0 < \alpha_g < \beta_g < +\infty$.

We are now in a position to apply Lemma 3.23, which asserts the existence of an interval $(a, b) \subset \mathbb{R}$ such that $f_t g$ is elliptic for all $t \in (a, b)$. Moreover there is a dense subset $A \subset (a, b)$ such that $f_t \in S$ for all $t \in A$. One possibility is there exists $t \in A$ such that $f_t g$ is an elliptic map of infinite order. In this case Theorem 3.24 tells us that $\langle f_t g, g \rangle$, and hence S , is dense in $\text{Aut}(\mathbb{H}^2)$. Otherwise, the collection of functions $\mathcal{F} = \{f_t g \mid t \in A\}$ are elliptic elements of finite order within S . By Lemma 3.23 the fixed points of \mathcal{F} accumulate

in \mathbb{H}^2 . Hence the group generated by \mathcal{F} is contained in S , and by Lemma 3.25 (i) this group is not elementary. Hence by Lemma 3.25 (ii) the group generated by \mathcal{F} contains an elliptic element of infinite order, h say. Now Theorem 3.24 tells us that $\langle h, g \rangle$ is dense in $\text{Aut}(\mathbb{H}^2)$, and so S is dense in $\text{Aut}(\mathbb{H}^2)$, as required. \square

Our next objective is to obtain a version of Theorem 3.2 for finitely-generated semigroups, as promised at the beginning of this chapter. However, before we do that, we must take a detour to study some anomalous semigroups that are exceptional cases in the version of Theorem 3.2 for finitely-generated semigroups and in Theorem 3.3.

5. Classification of finitely-generated semigroups and proof of Theorem 3.3

We now take a closer look at *exceptional* semigroups, given on page 10 in the introduction. Recall that a semigroup S of Möbius transformations is exceptional if it lies in $\text{Möb}(J)$ for some nontrivial closed interval J , and the collection of elements of S that fixes J as a set forms a nontrivial discrete group, and there is another element of S outside this discrete group that fixes one of the end points of J . One of the reasons that exceptional semigroups are special is that, although finitely-generated semigroups contained in $\text{Möb}(J)$ are usually semidiscrete, exceptional semigroups are never semidiscrete.

LEMMA 3.27. *Exceptional semigroups are not semidiscrete.*

PROOF. By conjugating, we may assume that $J = [0, +\infty]$ and the group part of S comprises maps of the form $g_n(z) = \lambda^n z$, $n \in \mathbb{Z}$, where $\lambda < 1$. Furthermore, we may assume that there is a map $f(z) = az + b$ in the inverse free part of S , where $a > 0$ and $b > 0$. Observe that, since $f \in S \setminus S^{-1}$, the maps

$$g_n f g_n^{-1}(z) = az + \lambda^n b$$

lie in $S \setminus S^{-1}$ for each $n \in \mathbb{N}$. Therefore the map $h(z) = az$ lies in the closure $\overline{S \setminus S^{-1}}$ of $S \setminus S^{-1}$ in $\text{Aut}(\mathbb{D})$. By composing h with the maps g_n we see that the identity belongs to $\overline{S \setminus S^{-1}}$, and so S is not semidiscrete. \square

In the remainder of this section we explain how it is straightforward to identify a finitely-generated exceptional semigroup by examining the generating set alone. We need the following lemma.

LEMMA 3.28. *Suppose that S is a semigroup contained in $\text{Möb}(J)$, for some nontrivial closed interval J , that is generated by a set \mathcal{F} . Then S has an element that fixes exactly one of the end points of J if and only if \mathcal{F} contains such an element.*

PROOF. If \mathcal{F} has an element that fixes exactly one of the end points of J , then certainly S does too, as $\mathcal{F} \subseteq S$. Conversely, suppose that no element of \mathcal{F} fixes exactly one of the end points of J . Suppose that an element f of S fixes one of the end points a of J . We can choose maps f_i in \mathcal{F} such that $f = f_1 \cdots f_n$. Observe that $f(J) \subseteq f_1(J)$, so $a \in f_1(J)$, which implies that $f_1(a) = a$. Therefore the element $f_2 \cdots f_n$ of S fixes a , and, by our assumption, $f_1(J) = J$. Repeating this argument we see that $f_j(J) = J$ for $j = 1, \dots, n$, so $f(J) = J$. We have now proved that no element of S fixes exactly one end point of J , as required. \square

Suppose now that we have a semigroup S generated by a finite set \mathcal{F} , and we wish to know whether S is exceptional. First look for a collection of two or more loxodromic transformations in \mathcal{F} with the same axis. Lemma 3.6 tells us precisely when this collection of loxodromic maps generates a discrete group. If there is no such collection, or more than one such collection, then S is not exceptional. Let us suppose there is indeed one such collection, with axis γ , and this collection does generate a discrete group. In order for S to be exceptional, every element of \mathcal{F} must lie in $\text{Möb}(J)$, where J is one of the two intervals in $\overline{\mathbb{R}}$ with the same end points as γ . If this is the case, then Lemma 3.28 tells us that S is exceptional if and only if \mathcal{F} contains an element that fixes exactly one of the end points of J .

5.1. A classification of finitely-generated semigroups. Here we prove Corollary 3.30, which is a version of Theorem 3.2 for finitely-generated semigroups. Given a closed interval J , let us denote by $\text{Möb}_0(J)$ the group part of $\text{Möb}(J)$, which comprises those elements in $\text{Möb}(J)$ that fix J as a set. It is the one-parameter family of loxodromic

Möbius transformations whose fixed points are the end points of J . If $J = [0, +\infty]$, then $\text{Möb}_0(J)$ consists of all maps of the form $z \mapsto \lambda z$, where $\lambda > 0$.

THEOREM 3.29. *Let S be a finitely-generated semigroup that is contained in $\text{Möb}(J)$ for some nontrivial closed interval J . Then S is semidiscrete if and only if it is nonexceptional and not dense in $\text{Möb}_0(J)$.*

PROOF. By Lemma 3.27, exceptional semigroups are not semidiscrete, and, obviously, nor are semigroups that are dense in $\text{Möb}_0(J)$. Suppose now that S is neither exceptional nor dense in $\text{Möb}_0(J)$. By conjugation, we may assume that $J = [0, +\infty]$. Let us denote by T the collection of maps in S that fix J as a set. Each element of T has the form $z \mapsto \lambda z$, where $\lambda > 0$. We split the remaining argument into two cases depending on whether (i) T is a nontrivial discrete group, or (ii) all elements $z \mapsto \lambda z$ of T satisfy $\lambda \geq 1$, or else they all satisfy $\lambda \leq 1$. We also include in (ii) the possibility that $T = \emptyset$. Let \mathcal{F} be a finite generating set for S .

In case (i), as S is not an exceptional semigroup, there is an interval $K = [a, b]$, where $0 < a < b < +\infty$, such that if $f \in \mathcal{F}$ and f does not fix J as a set, then $f(J) \subseteq K$. Suppose now that F_n is a sequence of distinct elements of S . We wish to show that F_n cannot converge to the identity, because doing so will demonstrate that S is semidiscrete. If $F_n \in T$ for infinitely many $n \in \mathbb{N}$, then certainly F_n cannot converge to the identity because T is discrete. Otherwise, when n is sufficiently large, we can write $F_n = G_n f_n H_n$, where $G_n \in T$ (possibly $G_n = I$), $f_n \in \mathcal{F} \setminus T$ and $H_n \in S$. Notice that

$$f_n H_n(J) \subseteq f_n(J) \subseteq K.$$

Let $G_n(z) = \lambda_n z$, where $\lambda_n > 0$. Either $\lambda_n \geq 1$, in which case $F_n(0) \geq f_n H_n(0) \geq a$, or $\lambda_n \leq 1$, in which case $F_n(\infty) \leq f_n H_n(\infty) \leq b$. Therefore F_n does not converge to the identity.

Let us move on to case (ii), and let us assume that every element of T has the form $z \mapsto \lambda z$, where $\lambda \leq 1$ (the other case is similar). We include the possibility that $T = \emptyset$ in our analysis. Let D denote the open top-right quadrant of the complex plane. This is the hyperbolic half-plane in \mathbb{H}^2 with boundary J . There exists a Euclidean line ℓ that

intersects \mathbb{R} at some point $x < 0$ and has a positive slope chosen such that it intersects D with the property that for any loxodromic element of \mathcal{F} that has $\alpha_f = \infty$ and $f(0) > 0$, we have $\beta_f < 0$, and for any element $f \in \mathcal{F} \setminus T$ that does not fix ∞ , the domain $f(D)$ lies entirely below ℓ . This is possible since \mathcal{F} is finite. Let U denote the subset of D consisting of points below the line ℓ .

It is easy to check that each element in $\mathcal{F} \setminus \{I\}$ maps U strictly inside itself. Since each $f \in \mathcal{F} \cap T$ maps ℓ to a parallel Euclidean line strictly below ℓ , we see each such f maps U strictly inside itself. If $f \in \mathcal{F} \setminus T$ and f does not fix ∞ , then f maps D to a Euclidean semicircle bounded by a hyperbolic geodesic, both of whose landing points lie in $[0, +\infty)$. We choose the slope of ℓ so that each such semicircle is contained in U . If $f \in \mathcal{F} \setminus T$ and $f(\infty) = \infty$, then possibly f is parabolic, in which case $f(z) = z + c$ for some $c \in (0, +\infty)$, and so f maps U strictly inside itself. Otherwise f is loxodromic, and either it has attracting fixed point ∞ and repelling fixed point in $(-\infty, 0)$, in which case by choice of x it can be seen that f maps U strictly inside itself; or it has repelling fixed point ∞ and attracting fixed point in $(0, +\infty)$, and the same conclusion can be drawn in this case too.

Now for each $f \in \mathcal{F} \setminus \{I\}$ we can choose a point $z_f \in \mathbb{H}^2 \setminus \overline{U}$ such that $f(z_f) \in U$. Any element of $S \setminus \{I\}$ maps at least one point of the finite set $\{z_f \mid f \in \mathcal{F} \setminus \{I\}\}$ into U . To see this, express the given element as a word in the generating set $\mathcal{F} \setminus \{I\}$ and consider the action of the element on z_f , where f is the rightmost term in this word. It follows that S is semidiscrete. \square

We can now state a version of Theorem 3.2 for finitely-generated semigroups, which follows immediately from Theorems 3.2 and 3.29.

COROLLARY 3.30. *Let S be a finitely-generated semigroup. Then S is either*

- (i) *elementary; or*
- (ii) *semidiscrete; or*
- (iii) *contained in $\text{Möb}(J)$, for some nontrivial closed interval J , and is either exceptional or dense in $\text{Möb}_0(J)$; or*
- (iv) *dense in $\text{Aut}(\mathbb{D})$.*

5.2. Proof of Theorem 3.3. Our next task is to prove Theorem 3.3. This theorem is the semigroups counterpart of a well-known result (see for example [3, Theorem 5.4.2]) in the theory of Fuchsian groups, which says that a nonelementary group of Möbius transformations is discrete if and only if every two-generator subgroup is discrete. Here is a version of that result for semigroups, which, unlike Theorem 3.3, does not assume that the semigroup is finitely-generated.

THEOREM 3.31. *Let S be a nonelementary semigroup that is not contained in $\text{Möb}(J)$, for any nontrivial closed interval J . Then S is semidiscrete if and only if every two-generator semigroup contained in S is semidiscrete.*

This theorem is an immediate corollary of Theorem 3.26. The assumption that S is not contained in $\text{Möb}(J)$ cannot be removed. To see why this is so, consider, for example, the semigroup S generated by the maps $(1 - 1/n)z$ for $n = 2, 3, \dots$. Clearly S is not semidiscrete, but every two-generator semigroup within S is semidiscrete. This particular semigroup is elementary, but it is easy to adjust it to give a nonelementary example: simply append to S any element of $\text{Möb}([0, +\infty])$ that fixes neither 0 nor ∞ .

Let us now turn to Theorem 3.3, which states that any finitely-generated nonexceptional semigroup S contained in $\text{Aut}(\mathbb{D})$ is semidiscrete if and only if every two-generator semigroup contained in S is semidiscrete.

PROOF OF THEOREM 3.3. If S is semidiscrete, then certainly any two-generator semigroup in S is semidiscrete. Conversely, suppose that every two-generator semigroup contained in S is semidiscrete. By Theorem 3.26, S is either elementary, semidiscrete or contained in $\text{Möb}(J)$, for some closed interval J . If S elementary and is not contained in the stabiliser group of z , for some $z \in \overline{\mathbb{C}}$, then one can check each case in Theorem 3.5 and verify that S semidiscrete. If S is contained in $\text{Möb}(J)$ and is not elementary or semidiscrete, then Corollary 3.30 tells us that S is dense in $\text{Möb}_0(J)$. However, Lemma 3.6(iii) shows that this cannot be so; therefore S cannot be contained in $\text{Möb}(J)$ but nonelementary or semidiscrete after all. We conclude that, in each case, S is semidiscrete. \square

6. Intersecting limit sets

In this section we prove Theorem 3.4, which says that if S is a nonelementary, finitely-generated and semidiscrete semigroup contained in $\text{Aut}(\mathbb{D})$, then $\Lambda^+(S) = \Lambda^-(S)$ if and only if S is a group. The next lemma is an important step in establishing this result.

LEMMA 3.32. *Let S be a nonelementary semigroup contained in $\text{Aut}(\mathbb{B}^3)$ that satisfies $\Lambda^-(S) \subseteq \Lambda^+(S)$. Let $f \in S$, $p \in \Lambda^-(S)$, and let U be an open neighbourhood of p . Then there exists an element g of S such that fg is loxodromic with attracting fixed point in U .*

PROOF. By Theorem 3.21, we can choose a loxodromic element h_1 of S such that $\beta_{h_1} \in U$. Now observe that for any point ζ in \mathbb{B}^3 , the sequence $(fh_1^n)^{-1}(\zeta)$ converges ideally to p . By replacing h_1 with a power of h_1 if necessary, we can assume that fh_1 is loxodromic, and using the observation that we just made we can assume that $\beta_{fh_1} \in U$.

Let V be an open neighbourhood of β_{fh_1} such that $fh_1(V) \subseteq U$. Appealing to Theorem 3.21 again, we can choose a loxodromic element h_2 of S such that $\alpha_{h_2} \in V$. By replacing h_2 by a power of h_2 if necessary we can ensure that, first, $h_2(U) \subseteq V$ and, second, fh_1h_2 is loxodromic. Since $fh_1h_2(U) \subseteq fh_1(V) \subseteq U$ we see that $\alpha_{fh_1h_2} \in U$. The lemma now follows on choosing $g = h_1h_2$. \square

The main theorem of this section follows.

THEOREM 3.33. *Let S be a finitely-generated, nonelementary semidiscrete subsemigroup of $\text{Aut}(\mathbb{D})$ that satisfies $\Lambda^-(S) \subseteq \Lambda^+(S)$. Then S is a group.*

PROOF. Let $g \in S$. Choose two distinct points u and v in $\Lambda^-(S)$. By Lemma 3.32 it is possible to construct a composition sequence $F_n = f_1 \cdots f_n$ generated by S that accumulates at both u and v . To see this, observe that given $F_n = f_1 \cdots f_n$ such that $F_n(0)$ is within ϵ (with respect to the chordal metric) of u , we can, by Lemma 3.32, choose $h \in S$ such that α_{F_nh} is within $\epsilon/2$ of v . Then for all large enough $k \in \mathbb{N}$, the point $(F_nh)^k(0)$ is within ϵ of v ; hence we can choose $f_{n+1} = h(F_nh)^{k-1}$ such that $F_{n+1}(0)$ is within ϵ of v . Furthermore, we can easily fashion this composition sequence such that $f_i = g$ for infinitely many positive integers i . Notice that F_n is discrete, because S is semidiscrete,

by Theorem 2.8.

If F_n is an escaping sequence, then S is a group, by Theorem 3.9. If the sequence F_n is not an escaping sequence then because it is discrete, it has a constant subsequence F_{n_1}, F_{n_2}, \dots . Choose $m > n_1$ such that $f_m = g$, and choose k such that $n_k < m \leq n_{k+1}$. Then

$$f_{n_k+1} f_{n_k+2} \cdots f_{n_{k+1}} = F_{n_k}^{-1} F_{n_{k+1}} = I.$$

Hence $g^{-1} \in S$. As we chose g arbitrarily from S , we conclude that S must be a group. \square

Theorem 3.4 given at the start of this chapter is an immediate consequence of this theorem. By referring to Section 1 of this chapter we see that the only finitely-generated semidiscrete semigroups that do not contain loxodromic elements are either finite groups of elliptic maps with a common fixed point, or semidiscrete semigroups of parabolic maps with a common fixed point. All these semigroups satisfy $\Lambda^- \subseteq \Lambda^+$, but some of the latter type are not groups, such as $\langle z \mapsto z + 1 \rangle$.

Theorem 3.33 only applies to subsemigroups of $\text{Aut}(\mathbb{D})$; however, provided that we additionally suppose the forward limit set of S is not connected, its proof can be adapted for subsemigroups of $\text{Aut}(\mathbb{B}^3)$. First we give a preparatory lemma.

LEMMA 3.34. *Suppose \mathcal{F} is a bounded collection of elements in $\text{Aut}(\mathbb{B}^3)$. If F_n is an escaping composition sequence generated by \mathcal{F} , then $\Lambda^+(F_n)$ is connected.*

PROOF. Suppose $F_n = f_1 \cdots f_n$ where f_n belongs to \mathcal{F} for each $n \in \mathbb{N}$. Choose any $\epsilon > 0$. Since \mathcal{F} is bounded in $\text{Aut}(\mathbb{B}^3)$, for any $\zeta \in \mathbb{B}^3$ there is some positive real number M such that $\rho(\zeta, f(\zeta)) \leq M$ for each $f \in \mathcal{F}$. Hence $\rho(F_n(\zeta), F_{n+1}(\zeta)) = \rho(\zeta, f_{n+1}(\zeta)) \leq M$ for all n . Since $F_n(\zeta)$ accumulates only on the ideal boundary and $\rho(F_n(\zeta), F_{n+1}(\zeta)) \leq M$, we have $\chi(F_n(\zeta), F_{n+1}(\zeta)) < \epsilon$ for all large enough n . Now suppose towards contradiction that $\Lambda^+(F_n)$ is not connected. Then we can choose two open subsets of \mathbb{S}^2 , U and V , both of which meet $\Lambda^+(F_n)$, that are a positive distance apart, and are such that $\Lambda^+(F_n) \subseteq U \cup V$. It follows that for all large enough n we have $\chi(F_n(\zeta), U \cup V) < \epsilon$. Hence can choose

positive integers p and q , where $p < q$, and such that both $\chi(F_p(\zeta), U)$ and $\chi(F_q(\zeta), V)$ are less than ϵ , and that satisfy $\chi(F_n(\zeta), F_{n+1}(\zeta)) < \epsilon$ and $\chi(F_n(\zeta), U \cup V) < \epsilon$ for each $n = p, p+1, \dots, q-1$. By choosing ϵ small enough, for example $\epsilon < \frac{1}{4}\chi(U, V)$, this contradicts the fact that $\chi(F_n(\zeta), U \cup V) < \epsilon$ for each $n = p, p+1, \dots, q-1$. \square

THEOREM 3.35. *Let S be a nonelementary, finitely-generated semidiscrete subsemigroup of $\text{Aut}(\mathbb{B}^3)$ that satisfies $\Lambda^-(S) \subseteq \Lambda^+(S)$. If $\Lambda^+(S)$ is not connected then S is a group.*

PROOF. Suppose towards contradiction that Λ^- is contained in one connected component of Λ^+ , which we shall denote by Λ_0^+ . Let $U \subseteq \mathbb{S}^2$ be any open set that meets Λ^+ . By Theorem 1.8(i) applied to S^{-1} we can choose a loxodromic transformation $g \in S$ with attracting fixed point in U . Since Λ^- is contained in Λ_0^+ , the repelling fixed point of g , namely β_g , must lie in Λ_0^+ . Since each element of S maps any component of Λ^+ within another component, it follows that g must map Λ_0^+ within itself. Since S is nonelementary, Λ^+ is a perfect set, so Λ_0^+ contains points other than β_g . Hence for all large enough n , the set $g^n(\Lambda_0^+)$ meets U . Since U was arbitrary, Λ_0^+ must be the only component of Λ^+ , contradicting our assumption that Λ^+ is not connected. It follows that we can assume Λ^- is not contained within a single connected component of Λ^+ .

Now just as in the proof of Theorem 3.33, for any pair of distinct points u and v in Λ^- , which we take to lie in different components of Λ^+ , we can find a composition sequence $F_n = f_1 \cdots f_n$ that accumulates at both points. (Strictly speaking, Theorem 3.33 assumes $S \subseteq \text{Aut}(\mathbb{D})$, but just the same argument works here.) Moreover, for any $g \in S$ we can choose F_n such that $f_i = g$ for infinitely many $i \in \mathbb{N}$. If F_n is an escaping sequence then by Lemma 3.34 $\Lambda^+(F_n)$ is connected, and so Λ^+ must have a connected component that contains u and v , which contradicts the fact that u and v lie in different components of Λ^+ . Hence F_n cannot be an escaping sequence, and because F_n is discrete it has a constant subsequence F_{n_1}, F_{n_2}, \dots . Choose $m > n_1$ such that $f_m = g$, and choose k such that $n_k < m \leq n_{k+1}$. Then

$$f_{n_k+1}f_{n_k+2} \cdots f_{n_{k+1}} = F_{n_k}^{-1}F_{n_{k+1}} = I.$$

Hence $g^{-1} \in S$. As we chose g arbitrarily from S , we conclude that S must be a group. \square

We suspect the assumption that $\Lambda^+(S)$ is not a connected set in Theorem 3.35 can be replaced with the assumption that $\Lambda^+(S) \neq \overline{\mathbb{C}}$. Whether or not this can be done is an open problem under investigation. Nevertheless, it can be shown that if there exists a nonelementary finitely-generated semidiscrete semigroup S whose forward and backward limit sets are equal to the connected set $\Lambda \neq \overline{\mathbb{C}}$, then S is contained in a Kleinian group with limit set Λ .

CHAPTER 4

Limit sets

1. Introduction

In this chapter we study the limit sets of a semigroup, their internal structure, and what limit sets can tell us about the semigroup that generates them. We have already utilised analogues of the conical and horospherical limit sets from the theory of Kleinian groups in Chapter 3. In Section 2 we introduce further noteworthy subsets of the limit sets and study their properties. In Section 3 we study semigroups whose limit sets are disjoint, and expand on recent work of Fried, Marotta and Stankewitz [15]. Finally, we consider how the limit sets of a semigroup behave under perturbation of its generators. Most of the results in this chapter relate to semigroups in any number of dimensions, but are not as strong as those obtained in Chapter 3 on semigroups acting on two dimensions.

We next give two lemmas from [15], which for completeness we record here.

LEMMA 4.1. (*Topological transitivity*)

Let S be a semigroup and suppose $U \subseteq \mathbb{S}^2$ is any open set that meets Λ^- . Then $S(U)$ is either the whole ideal boundary, or the whole ideal boundary missing one point. In the latter case, the missing point is fixed by each element in S , and in particular S is elementary.

PROOF. First note that $S(U)$ is forward invariant with respect to S . By Theorem 1.8 (ii) we see that S is not a normal family on U . Hence by Montel's theorem, $S(U)$ misses at most two points on $\overline{\mathbb{C}}$. First suppose towards contradiction that $S(U)$ misses exactly two points, say w and z . Then since $S(U)$ is forward invariant, its complement, $\{w, z\}$, is backward invariant. This means that $\{w, z\}$ is a finite orbit of S , and in particular, S is elementary. By consulting Proposition 1.7 and its proceeding discussion, we see that S belongs either to case (c) or to case (f) given in the Proposition. By checking these

possibilities it can be seen that in fact $S(U)$ can miss at most one point, contrary to our assumption. Hence $S(U)$ misses at most one point in \mathbb{S}^2 , z say. Then it follows that S fixes $\{z\}$, as the complement of the forward invariant set $S(U)$. \square

LEMMA 4.2. *Suppose S is a semigroup and z is a point in Λ^- that is not fixed by every element in S . Then $\Lambda^- = \overline{S^{-1}(z)}$. Similarly if $z \in \Lambda^+$ is not fixed by every element in S , then $\Lambda^+ = \overline{S(z)}$.*

PROOF. Pick any $\zeta \in \Lambda^-$, and let U be any open neighbourhood of ζ . Then by Theorem 4.1 we have $z \in g(U)$ for some $g \in S$, that is $g^{-1}(z) \in U$. Since U was arbitrary, ζ is an accumulation point of $S^{-1}(z)$, and so $\Lambda^- \subseteq \overline{S^{-1}(z)}$. Conversely, since $z \in \Lambda^-$ and Λ^- is backward invariant, we see that $S^{-1}(z) \subseteq \Lambda^-$. Taking closures we obtain the reverse inclusion. Applying the same argument to S^{-1} gives the second statement. \square

As we are studying limit sets it is convenient to introduce notation to represent their complements. Accordingly, for a semigroup S we define $\Omega^+(S) = \mathbb{S}^2 \setminus \Lambda^+(S)$ and $\Omega^-(S) = \mathbb{S}^2 \setminus \Lambda^-(S)$. As usual, if there is no danger of ambiguity we suppress mention of S .

2. On the structure of the limit sets

2.1. The sets D^+ and D^- . Here we introduce two important subsets of the limit sets of a semigroup. For any semigroup S we define the sets

$$D^-(S) = \left\{ z \in \Lambda^-(S) \mid \Lambda^-(S) \subseteq \overline{S(z)} \right\}$$

and

$$D^+(S) = \left\{ z \in \Lambda^+(S) \mid \Lambda^+(S) \subseteq \overline{S^{-1}(z)} \right\}.$$

As usual we shall simply write D^+ or D^- if the semigroup is clear from the context. We first note that D^+ is a forward invariant set and D^- is backward invariant. When S is a nonelementary group, the sets D^+ or D^- are both equal to $\Lambda^+ = \Lambda^-$, and so any utility of these sets can be realised only when S is not a group – there is no useful analogue of D^+ or D^- in the theory of Kleinian groups.

By Lemma 4.2 for any $z \in \Lambda^+$, the closure of the orbit of $S(z)$ is equal to Λ^+ . If moreover z lies in D^+ , then the closure of the orbit of $S^{-1}(z)$ covers Λ^- . Recall that if f is a rational

function acting on $\overline{\mathbb{C}}$ then the Julia set $J(f)$ is defined as the set of non-normality of the family $\{f^n \mid n \in \mathbb{N}\}$. The following result (sometimes called topological transitivity) is well known: for any rational function f the set of points $z \in J(f)$ for which the orbit of z is dense in $J(f)$, is itself dense in $J(f)$. The following theorem is motivated by the proof of that result, as given in [32, Corollary 4.16].

THEOREM 4.3. *Suppose S is a semigroup such that no point in Λ^- is fixed by every element in S . Then D^- is a dense G_δ subset of Λ^- with respect to the subspace topology on Λ^- .*

PROOF. Working in the chordal metric, let $\{U_{j,k} \mid k = 1, \dots, n(j)\}$ be an open cover of Λ^- , where $U_{j,k}$ has chordal diameter less than $1/j$ and $U_{j,k}$ meets Λ^- for each $j \in \mathbb{N}$ and each $k = 1, \dots, n(j)$. Let $V_{j,k} = S^{-1}(U_{j,k})$ for each $j \in \mathbb{N}$ and $k = 1, \dots, n(j)$. Since $U_{j,k}$ meets Λ^- and (by Lemma 4.2) $\Lambda^- = \overline{S^{-1}(z)}$ for all $z \in \Lambda^-$, we can pick $z \in U_{j,k} \cap \Lambda^-$ so that as

$$S^{-1}(z) \subseteq S^{-1}(U_{j,k})$$

it follows that

$$\overline{S^{-1}(z) \cap \Lambda^-} \subseteq \overline{S^{-1}(U_{j,k}) \cap \Lambda^-},$$

and since $S^{-1}(z)$ is a dense subset of Λ^- we have

$$\Lambda^- \subseteq \overline{S^{-1}(U_{j,k}) \cap \Lambda^-}.$$

Hence $V_{j,k} \cap \Lambda^-$ is dense in Λ^- for each $j \in \mathbb{N}$ and $k = 1, \dots, n(j)$. As a complete metric space, Λ^- is a Baire space, that is, any countable intersection of open dense subsets of Λ^- is dense in Λ^- . It follows that

$$D = \bigcap_{\substack{j \in \mathbb{N} \\ k \leq n(j)}} (V_{j,k} \cap \Lambda^-) = \Lambda^- \cap \bigcap_{\substack{j \in \mathbb{N} \\ k \leq n(j)}} V_{j,k}$$

is dense in Λ^- . Now if $z \in D$ then for all $j \in \mathbb{N}$ and $k \in \{1, \dots, n(j)\}$ there is some $g \in S$ such that $g(z) \in U_{j,k}$. Hence $S(z)$ has points arbitrarily close to any point in Λ^- , and so $\overline{S(z)}$ covers Λ^- . This proves that D is a dense G_δ set with respect to the subspace topology on Λ^- , and that $D \subseteq D^-$.

By unpacking the definition, we verify that $D^- \subseteq D$: For suppose $z \in D^-$; since each open set $U_{j,k}$ meets Λ^- and $S(z)$ is dense in Λ^- , $S(z)$ meets each $U_{j,k}$. In other words there exists some $g \in S$ such that $z \in g^{-1}(U_{j,k})$, so that $z \in S^{-1}(U_{j,k})$. Since this holds for each $U_{j,k}$ and z lies in Λ^- , it follows that z lies in D^- by definition. \square

COROLLARY 4.4. *The set D^- meets Λ^+ if and only if $\Lambda^- \subseteq \Lambda^+$.*

PROOF. Clearly if $\Lambda^- \subseteq \Lambda^+$ then D^- meets Λ^+ . Conversely suppose z lies in $D^- \cap \Lambda^+$; then we have

$$\Lambda^- \subseteq \overline{S(z)} = \Lambda^+,$$

as required. \square

Recall that a subset of a topological space is meagre if it is the countable union of nowhere dense sets. A residual set is the complement of a meagre set. Meagre sets can be regarded as topologically small sets – the family of meagre sets form a σ -ideal, that is the family contains all subsets and all countable unions of its elements. In a complete metric space such as Λ^- , a set is residual if and only if it contains a dense G_δ set. This means we can think of D^- as a topologically large set, at least with respect to the subspace topology on Λ^- . It follows from Theorem 4.3 and Corollary 4.4 that unless Λ^- is contained in Λ^+ , then $\Lambda^+ \cap \Lambda^-$ is a meagre subset of Λ^- . Similarly, unless Λ^+ is contained in Λ^- , then $\Lambda^- \cap \Lambda^+$ is a meagre subset of Λ^+ . Recall that in dimension two, by Theorem 3.33, if S is finitely-generated, nonelementary, semidiscrete and either limit set is contained in the other, then S is a group. Hence we can restate Theorem 3.33 as follows: a finitely-generated, nonelementary and semidiscrete subsemigroup of $\text{Aut}(\mathbb{D})$ is a group if and only if D^+ meets D^- , in which case $D^+ = D^-$.

Figure 4.1 shows the limit sets of the semigroup S generated by the maps $f(z) = \frac{a}{1+z}$, $g(z) = \frac{a-1+2ia^{1/2}}{1+z}$ and $h(z) = \frac{-0.25}{1+z}$, where $a = -0.1 + 0.7i$. These maps are chosen such that f , gh and hg are parabolic elements. The forward limit set is shown in red, while the backward limit set is shown in blue. The intersection $\Lambda^+(S) \cap \Lambda^-(S)$ is the set of fixed points of f , gh and hg , which is certainly a meagre subset of both $\Lambda^+(S)$ and $\Lambda^-(S)$.

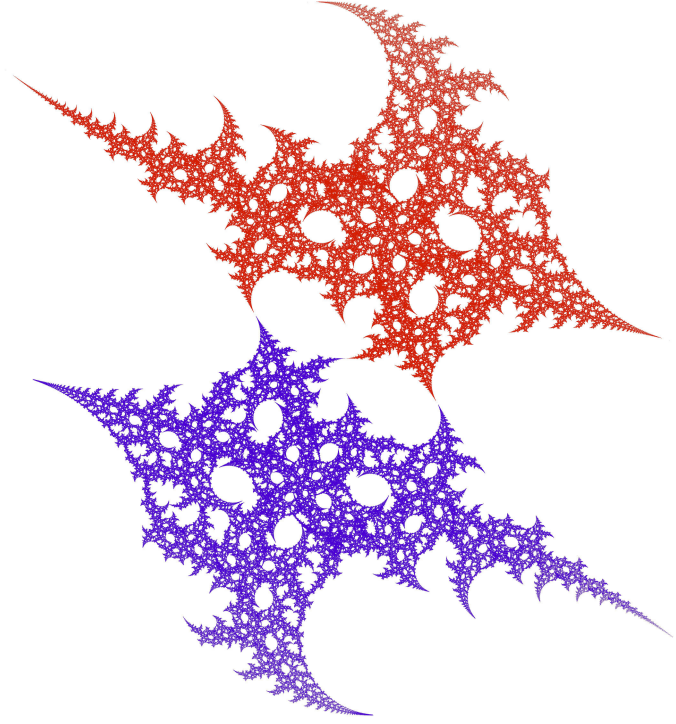


FIGURE 4.1. Limit sets of a semigroup whose intersection is finite

The next theorem shows that if $\Lambda^- \not\subseteq \Lambda^+$, then $D^- \subseteq \Lambda_c^-$. We know that if $\Lambda^- \not\subseteq \Lambda^+$ then D^- does not meet Λ^+ ; however there are many such semigroups for which Λ_c^- meets Λ^+ . For example, take a Schottky subgroup G of $\text{Aut}(\mathbb{D})$ and choose two disjoint closed discs I^- and I^+ that are orthogonal to the unit disc and that do not meet the Dirichlet region of G centred at 0. Choose an element f in $\text{Aut}(\mathbb{D})$ that maps the interior of the complement of I^+ onto I^- . The semigroup S generated by $G \cup \{f\}$ is semidiscrete, and has group part G strictly contained in S . As Schottky groups are convex-cocompact, that is the (hyperbolic) convex hull of the limit set intersected with any Dirichlet region is a bounded set with respect to the hyperbolic metric, the conical limit set of G is the limit set of G . Hence $\Lambda_c^-(S)$ meets $\Lambda^+(S)$ on at least the limit set of G , however $\Lambda^-(S) \not\subseteq \Lambda^+(S)$.

THEOREM 4.5. *Suppose S is a semigroup such that $\Lambda^- \not\subseteq \Lambda^+$; then $D^- \subseteq \Lambda_c^-$.*

PROOF. It follows from Corollary 4.4 that $D^- \cap \Lambda^+ = \emptyset$. Now choose $z \in D^-$, so by definition of D^- there exists a sequence $g_n \in S$ such that $g_n(z)$ converges to z . It follows that $z \in \Lambda_c^-(g_n)$ for otherwise $g_n(z)$ must accumulate exactly where $g_n(\zeta)$ accumulates for any point ζ in hyperbolic space. This set of accumulation points is contained in Λ^+ , and so cannot include z as D^- is disjoint from Λ^+ . Hence z is a backwards conical limit point of the sequence g_n , and so lies in the backward conical limit set of S . \square

It follows from the theorem above that provided $\Lambda^- \not\subseteq \Lambda^+$, then $\Lambda^- \setminus \Lambda_c^-$ is a meagre subset of Λ^- . Notice that this holds whether or not S is discrete or semidiscrete, and echoes the situation for a geometrically finite discrete group G , for which $\Lambda(G) \setminus \Lambda_c(G)$ is a countable set.

The semigroup of Theorem 3.11 is an example of a semigroup for which D^+ is not equal to Λ_c^+ , as we now explain. Consider (see Figure 4.2 below) the semigroup $S = \langle g, h \rangle$ where g, h are loxodromic maps such that $\alpha_g = \beta_h$ but β_g and α_h are distinct points.

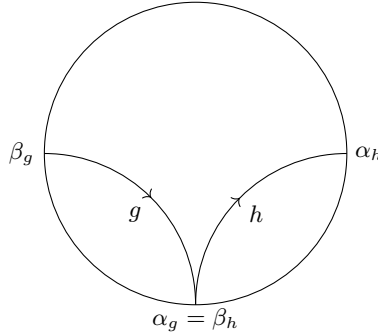


FIGURE 4.2. The axes of g and h

Then the closed interval contained in the unit circle that is separated from the axis of h by the axis of g is equal to the backward conical limit set. Similarly the closed interval contained in the unit circle that is separated from the axis of g by the axis of h is equal to the forward conical limit set. Hence the point $\alpha_g = \beta_h$ lies in both conical limit sets, and Λ^+ is the proper closed interval $[\beta_h, \alpha_h]$. As S fixes $\alpha_g = \beta_h$, this point cannot lie in either D^+ or D^- , because the S^{-1} orbit of $\alpha_g = \beta_h$ cannot accumulate everywhere in Λ^+ .

Similarly $\alpha_h \notin D^+$ because $\alpha_h \in \overline{\Omega^+}$ and $\overline{\Omega^+}$ is a backward invariant set, so that $S^{-1}(\alpha_h)$ cannot accumulate in the interior of Λ^+ .

Before introducing the next interesting subset of Λ^+ , we give an explicit description of $D^+(S)$ when S is the Cantor semigroup, that is $S = \langle f_0, f_2 \rangle$ where $f_i(z) = \frac{1}{3}(z + i)$ for $i = 0, 2$. We shall shortly (in subsection 2.3) show that $D^+(S)$ is the set of points in $\Lambda^+(S)$ whose base-3 expansion contains every finite string in $\{0, 2\}$. Since the limit sets of S are disjoint the forward conical limit is the full limit set, and so $D^+(S)$ is strictly contained in $\Lambda_c^+(S)$ in this case.

2.2. The point transitive limit set and an inclusion theorem. The *forward point transitive limit set* of S , which we shall denote by $\Lambda_p^+(S)$, is the set of points z on the ideal boundary with the following property. For any ζ in hyperbolic space and half-line γ landing at z , there exists an escaping sequence $g_n \in S$ such that $\rho(g_n(\zeta), \gamma) \rightarrow 0$ as $n \rightarrow \infty$. It is easy to see that $\Lambda_p^+(S)$ is forward invariant, and so for nonelementary semigroups $\Lambda_p^+(S)$ is either empty or dense in Λ^+ . The *backward point transitive limit set*, denoted by $\Lambda_p^-(S)$, is defined as $\Lambda_p^+(S^{-1})$. We remark that Λ_p^+ , and indeed Λ_c^+ , are $G_{\delta\sigma}$ sets, as can be seen from the set-theoretic descriptions of Λ_p^+ and Λ_c^+ given, for example, in Nicholls [34, Theorem 2.4.3]. The point transitive limit sets of Kleinian groups are utilised in number theory, see for example Lehner [25, Chapter 10]. Here we make little use of them; however they have a pleasing geometric definition and we can relate them to the other sets introduced thus far.

THEOREM 4.6. *Suppose S is a semigroup such that Ω^- is not empty; then $\Lambda_p^+ \subseteq D^+$.*

PROOF. Suppose $z \in \Lambda_p^+$ so that for all $\zeta \in \mathbb{B}^3$ there exists an escaping sequence g_n in S such that

$$\rho(g_n(\zeta), \gamma) \rightarrow 0 \text{ as } n \rightarrow \infty,$$

where γ is the geodesic half-ray starting at $\mathbf{0}$ and landing at z . It follows that

$$\rho(\zeta, g_n^{-1}(\gamma)) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Now for any point $w \in \Omega^-$ we can choose ζ to be an arbitrarily small Euclidean distance away from w compared to the distance from w to Λ^- . It follows from the geometry that

there is some $K > 0$ (depending on the distance from w to Λ^-) such that one of the end points of $g_n^{-1}(\gamma)$ accumulates at most a distance $K(1 - |\zeta|)$ away from w . By choosing ζ close enough to w , any point on the boundary within $K(1 - |\zeta|)$ of w lies inside Ω^- , and so it must be $g_n^{-1}(z)$ (and not $g_n^{-1}(\zeta)$) that accumulates within $K(1 - |\zeta|)$ of w . Since we can choose $1 - |\zeta|$ to be arbitrary small, it follows that $S^{-1}(z)$ accumulates at w itself. As w was arbitrary we have $\Omega^- \subseteq \overline{S^{-1}(z)}$, and so $\overline{\Omega^-} \subseteq \overline{S^{-1}(z)}$. Since Λ^+ is contained in $\overline{\Omega^-}$ the result follows. \square

For any $z \in D^+$, the closure of $S^{-1}(z)$ covers Λ^+ . The proof above shows that for $z \in \Lambda_p^+$, the closure of $S^{-1}(z)$ covers $\overline{\Omega^-}$.

Let μ be the Lebesgue measure on \mathbb{S}^2 , normalised so that $\mu(\mathbb{S}^2) = 1$. For Kleinian groups, the Lebesgue measure of the conical and point transitive limit sets are equal [34, Theorem 2.4.3]. For semigroups, provided $\Lambda^+ \not\subseteq \Lambda^-$, the sets D^+ and Λ_p^+ either have the same measure as Λ_c^+ , or have measure zero. In fact this is true of any forward invariant subset of Λ_c^+ . At its heart, the proof given in [34, Theorem 2.4.3] uses the Lebesgue density theorem, which can be found in [13, Equation 2.20], and we record here.

THEOREM 4.7. *Lebesgue density theorem* For $x \in \mathbb{S}^2$ let $B(x, r)$ be the open chordal ball of radius r centred at x , and let μ be the Lebesgue measure on \mathbb{S}^2 . If E is a measurable subset of \mathbb{S}^2 , then for μ -almost all $x \in E$ we have

$$\lim_{r \rightarrow 0} \frac{\mu(B(x, r) \cap E)}{\mu(B(x, r))} = \begin{cases} 1 & \text{if } x \in E \\ 0 & \text{if } x \notin E. \end{cases}$$

Our analogous result uses hyperbolic harmonic measure, a tool familiar to Kleinian group theorists. The hyperbolic harmonic measure of a measurable set $X \subseteq \mathbb{S}^2$ is the function

$$\omega_X : \mathbb{B}^3 \rightarrow \mathbb{R}$$

defined by

$$\omega_X(z) = \int_X \left(\frac{1 - |z|^2}{|z - \zeta|^2} \right)^2 d\mu(\zeta).$$

See [34, Chapter 5] for more on hyperbolic harmonic measure.

LEMMA 4.8. *Let S be a semigroup and let X be any measurable forward invariant subset of \mathbb{S}^2 . If $\mu(\Lambda_c^- \cap X) > 0$ then $\mu(X) = 1$.*

PROOF. For $z \in \mathbb{B}^3$ let $\omega_X(z)$ denote the hyperbolic harmonic measure of X on \mathbb{S}^2 . Then for any $g \in S$ we have

$$\omega_X(\mathbf{0}) \geq \omega_{g(X)}(\mathbf{0}) = \omega_X(g^{-1}(\mathbf{0})).$$

The inequality follows from $g(X) \subseteq X$ and the definition of ω_X , while the equality follows from the invariance property $\omega_{g(X)}(z) = \omega_X(g^{-1}(z))$ for all $g \in \text{Aut}(\mathbb{B}^3)$, given in [34, Equation 5.1.3]. The *conical limit* of $\omega_X(z)$ at a point w on the ideal boundary, is the limit, if it exists, of $\omega_X(z)$ as z converges to w from a bounded hyperbolic distance of some geodesic landing at w . Fatou's theorem [34, Theorem 5.1.5] tells us that the conical limit of $\omega_X(z)$ exists and agrees with the characteristic function of X at μ -almost every point on \mathbb{S}^2 . Since $\mu(\Lambda_c^- \cap X) > 0$ there exists some $w \in \Lambda_c^- \cap X$ such that $\omega_X(z)$ has limit 1 as z converges to w conically. Let us choose some sequence $g_n \in S$ such that $g_n^{-1}(\mathbf{0}) \rightarrow w$ conically. The above inequality gives

$$\omega_X(\mathbf{0}) \geq \omega_X(g_n^{-1}(\mathbf{0})) \rightarrow 1 \text{ as } n \rightarrow \infty.$$

But from the definition of hyperbolic harmonic measure, $\omega_X(\mathbf{0}) = \mu(X)$. Hence $\mu(X) = 1$. \square

COROLLARY 4.9. *If S is a nonelementary semigroup such that $\Lambda^+ \not\subseteq \Lambda^-$, then*

$$\Lambda_p^+ \subseteq D^+ \subseteq \Lambda_c^+.$$

Moreover, either $\mu(\Lambda_p^+) = \mu(\Lambda_c^+)$ or $\mu(\Lambda_p^+) = 0$. Similarly for D^+ , either $\mu(D^+) = \mu(\Lambda_c^+)$ or $\mu(D^+) = 0$.

PROOF. The inclusions follows from Theorems 4.5 and 4.6 noting that $\Lambda^+ \not\subseteq \Lambda^-$ implies Ω^- must be nonempty. As we remarked in the introduction to this subsection, both Λ_p^+ and Λ_c^+ are $G_{\delta\sigma}$ sets and so are certainly measurable, while that D^+ is measurable can be seen from its definition. Now suppose Λ_c^+ and Λ_p^+ do not have the same measure, that is $\mu(\Lambda_c^+ \cap X) > 0$ where X is the backward invariant set $\mathbb{S}^2 \setminus \Lambda_p^+$. Now by Lemma 4.8

(applied to S^{-1}) it follows that $\mu(X) = 1$, and so $\mu(\Lambda_p^+) = 0$. The argument for D^+ is identical. \square

We remark that in the case where S is discrete, the argument given in the proof of [34, Theorem 2.4.3] can be directly applied to S , and we can infer that the sets Λ_p^+ and Λ_c^+ have the same Lebesgue measure.

2.3. The buried points of the limit sets. For a semigroup S we define $B^+(S)$ to be the set of *buried* (sometimes called inaccessible) points in $\Lambda^+(S)$, that is the set of points in $\Lambda^+(S)$ that do not lie on the boundary of any connected component of $\Omega^+(S)$. As usual we abbreviate $B^+(S)$ to B^+ whenever the semigroup in question is clear. Because of the topological nature of this definition, the ambient space upon which S is taken to act is important. For example if S is the Cantor semigroup whose forward limit set is the middle-thirds Cantor set, then, regarded as a subset of the real line, only the countably many end points of the deleted intervals are not buried point of the limit set. However as a subset of the complex plane, no limit point is buried. Because of the difficulties in studying higher dimensions, we limit our study of buried points to semigroups acting on 2-dimensional hyperbolic space.

LEMMA 4.10. *If S is a nonelementary semigroup acting on $\text{Aut}(\mathbb{D})$, then B^+ is a forward invariant, dense G_δ subset of Λ^+ .*

PROOF. Let \mathcal{C} denote the set of components of Ω^+ , so that

$$\Lambda^+ = \mathbb{S}^2 \setminus \bigcup_{C \in \mathcal{C}} C$$

and

$$B^+ = \mathbb{S}^2 \setminus \bigcup_{C \in \mathcal{C}} \overline{C} = \bigcap_{C \in \mathcal{C}} \mathbb{S}^2 \setminus \overline{C} = \bigcap_{C \in \mathcal{C}} \Lambda^+ \setminus \overline{C}.$$

Hence B^+ is a G_δ subset of Λ^+ . Each C is an open interval, the closure of which meets the perfect set Λ^+ only at its end points, and so $\Lambda^+ \setminus \overline{C}$ is an open, dense subset of Λ^+ . As Λ^+ is a Baire space it follows from the Baire category theorem that B^+ is a dense subset of Λ^+ . It remains to show that B^+ is forward invariant. Given any $z \in B^+$ and $g \in S$, suppose towards contradiction that $g(z)$ is not a buried point. Then $g(z) \in \partial C$ for some

$C \in \mathcal{C}$, and so $z \in g^{-1}(\partial C)$. Since Ω^+ is backward invariant, g^{-1} maps each connected component inside another, so that $g^{-1}(\partial C) \subseteq \overline{C_1}$ for some $C_1 \in \mathcal{C}$. Since z lies in the complement of Ω^+ we must have $z \in \partial C_1$, contradicting that z is a buried point. This contradiction shows that $g(z)$ must be a buried point after all. \square

Recall that D^+ is also a forward invariant, dense G_δ subset of Λ^+ and so B^+ has some of the same dynamical and topological properties as D^+ . In two dimensions, one is contained in the other.

PROPOSITION 4.11. *If S is a nonelementary subsemigroup of $\text{Aut}(\mathbb{D})$ and $\Lambda^+ \not\subseteq \Lambda^-$, then $D^+ \subseteq B^+$.*

PROOF. We first consider the case where Λ^+ has empty interior, so that Λ^+ is a Cantor set. We can write the set of connected components of Ω^+ as

$$\{(p_i, q_i) \mid i \in I\}$$

where I is some countable index set and $(p_i, q_i) \subseteq \mathbb{S}^1$ is the open interval contained in the unit circle traversed by moving in an anticlockwise direction from p_i to q_i . We write $P = \{p_i \mid i \in I\}$ and $Q = \{q_i \mid i \in I\}$, so that $\Lambda^+ \setminus B^+ = P \cup Q$. Notice that since Λ^+ has no isolated points, P and Q are disjoint. Suppose there exists some z that lies in both $\Lambda^+ \setminus B^+$ and D^+ . Further suppose $z \in P$ and write $z = p_j$; we shall show that Λ^- covers Q . Since $z \in D^+$, $S^{-1}(z)$ accumulates at every point in Λ^+ and in particular at every point in Q . This means that for any $q \in Q$ there exists a sequence $g_n \in S$ such that

$$\lim_{n \rightarrow \infty} g_n^{-1}(p_j) = q.$$

Recall that g^{-1} is an orientation-preserving self-map of \mathbb{S}^1 that maps components of Ω^+ within components of Ω^+ . Hence there is some sequence $k(n)$ in I such that

$$g_n^{-1}(p_j, q_j) \subseteq (p_{k(n)}, q_{k(n)}).$$

One possibility is that there exists some subsequence of $k(n)$ whose elements are distinct. In this case we can assume without loss of generality that each $k(n)$ is distinct. Hence the diameters of $(p_{k(n)}, q_{k(n)})$ tend to zero as n tends to infinity, and since $g_n^{-1}(p_j)$ converges

to q , so does $g_n^{-1}(\zeta)$ for every $\zeta \in (p_j, q_j)$. Hence g_n^{-1} converges ideally to q , and so $q \in \Lambda^-$.

Otherwise there is some constant subsequence of $k(n)$, and in this case, without loss of generality, we can assume the sequence itself is constant. Hence

$$g_n^{-1}(p_j, q_j) \subseteq (p_k, q_k)$$

for some $k \in I$ and all n . Since $g_n^{-1}(p_j)$ converges to a point in Q and $P \cap Q = \emptyset$, we must have $q = q_k$. Again using that g_n^{-1} preserves orientation, $g_n^{-1}(\zeta)$ converges to q_k for every point ζ in (p_j, q_j) . This means that g_n^{-1} converges ideally to q_k , and so $q = q_k \in \Lambda^-$. By symmetry of argument if z lies in Q and D^+ , then Λ^- covers P . Hence the existence of any point z in both $\Lambda^+ \setminus B^+$ and D^+ implies that Λ^- covers at least one of the sets P or Q .

We claim that both P and Q are dense in Λ^+ . Then if $\Lambda^+ \setminus B^+$ and D^+ meet, then $P \subseteq \Lambda^-$ or $Q \subseteq \Lambda^-$; and so by taking closures we obtain $\Lambda^+ \subseteq \Lambda^-$, contrary to hypothesis. It follows from this contradiction that no such z exists and so $\Lambda^+ \setminus B^+ \subseteq \Lambda^+ \setminus D^+$, that is, $D^+ \subseteq B^+$ as required.

It now suffices to prove the claim. Each point in P is the limit of points in Q , that is $P \subseteq \overline{Q}$, and so $\overline{P} \subseteq \overline{Q}$. Similarly $\overline{Q} \subseteq \overline{P}$ and so $\overline{Q} = \overline{P}$, and it follows that

$$\Lambda^+ = \overline{P \cup Q} = \overline{P} \cup \overline{Q} = \overline{P} = \overline{Q}.$$

This completes the proof of the claim and proves the result in the case where Λ^+ has empty interior.

Now suppose Λ^+ has nonempty interior. If Λ^+ is equal to \mathbb{S}^1 , then B^+ is also equal to \mathbb{S}^1 ; hence $D^+ \subseteq B^+$. Otherwise we can choose a point z in $\Lambda^+ \setminus B^+$. In this case $\overline{\Omega^+}$ is a backward invariant set that does not cover Λ^+ . Since z lies in Λ^+ it lies in the closure of Ω^+ , but by the definition of B^+ it cannot lie in the interior of Ω^+ , and so $z \in \partial\Omega^+$. Since $\overline{\Omega^+}$ is backward invariant it follows that $S^{-1}(z)$ does not meet the complement of $\overline{\Omega^+}$, that is, the interior of Λ^+ . Hence z cannot lie in D^+ , which shows that $\Lambda^+ \setminus B^+ \subseteq \Lambda^+ \setminus D^+$; in other words $D^+ \subseteq B^+$, as required. \square

Notice that in the case where Λ^+ has nonempty interior, we did not use the fact that $S \subseteq \text{Aut}(\mathbb{D})$. This means that in all dimensions $D^+ \subseteq B^+$ whenever Λ^+ has a nonempty interior.

We ask whether or not the reverse inclusion of Proposition 4.11 holds, that is, do we have $B^+ \subseteq D^+$ under suitable hypotheses? The answer to this question is no, as can be seen by describing D^+ and B^+ explicitly in the case of a very well behaved and tractable semigroup, the Cantor semigroup, which we recall is generated by the maps $f_i(z) = \frac{1}{3}(z + i)$ for $i = 0, 2$. We remark that although the Cantor semigroup is an elementary semigroup, the following descriptions of D^+ and B^+ can be seen to generalise; for example to sub-semigroups of $\text{Aut}(\mathbb{D})$ that have disjoint limit sets and are generated by a finite set \mathcal{F} , with the property that $f(\Lambda^+) \cap g(\Lambda^+) = \emptyset$ for all distinct $f, g \in \mathcal{F}$. This means that the reverse inclusion of Proposition 4.11 fails for a large class of finitely-generated nonelementary semigroups.

Let $S = \langle f_0, f_2 \rangle$, the Cantor semigroup. Each point z in Λ^+ has a unique base-3 expansion

$$z = 0.i_1i_2 \dots i_n \dots$$

where each i_n lies in $\{0, 2\}$. We consider words, both infinite and finite, whose letters lie in $\{0, 2\}$. It is clear that B^+ is the set of those points $z = 0.i_1i_2 \dots i_n \dots$ in Λ^+ such that i_n is not eventually equal to 0 or eventually equal to 2. Since the forward and backward limit sets of S are disjoint we have $\Lambda_c^+ = \Lambda^+$. On the other hand, D^+ is the set of points in Λ^+ whose base-3 expansion contains every finite string in the letters $\{0, 2\}$ as a sub-string. To see this, first suppose $i_1i_2 \dots i_n \dots$ is an infinite word in $\{0, 2\}$ with this property. Then for any finite word w there are natural numbers $m < n$ such that $w = i_m \dots i_n$. Hence $f_{i_{m-1}}^{-1} \dots f_{i_1}^{-1}(z)$ is a point in Λ^+ whose base-3 expansion begins with w . Since w was arbitrary, $S^{-1}(z)$ accumulates at every point in Λ^+ , and so $z \in D^+$ by definition. Conversely suppose $z \in \Lambda^+$ has base-3 expansion $0.i_1i_2 \dots i_n \dots$. Suppose that some finite word, $w = w_1 \dots w_m$ say, does not feature anywhere in the infinite word $i_1i_2 \dots i_n \dots$. We claim that $S^{-1}(z)$ does not meet $f_{w_1} \dots f_{w_m}([0, 1])$. The claim follows since for any infinite

word $j_1 j_2 \dots j_n \dots$ and any $p \in \mathbb{N}$ we have

$$0.j_1 j_2 \dots j_n \dots \in f_{j_1} \dots f_{j_p}([0, 1]),$$

and that $f_{j_p}^{-1} \dots f_{j_1}^{-1}(0.i_1 i_2 \dots i_n \dots) \in (0, 1)$ if and only if $j_q = i_q$ for every $q = 1, 2, \dots, p$. Hence by the claim, $S^{-1}(z)$ avoids the interior of $f_{w_1} \dots f_{w_m}([0, 1])$ and so is not dense in Λ^+ . This implies that z cannot lie in D^+ .

It is interesting to consider how the set of attracting fixed points fits in this picture. It turns out that every attracting fixed point of S , except those of the generators f_0 and f_2 , lies in B^+ , but no attracting fixed points lie in D^+ . Indeed, if $i \in \{0, 2\}^n$ then the attracting fixed point of $F_i = f_{i_1} \dots f_{i_n}$ has base-3 expansion $0.i_1 \dots i_n i_1 \dots i_n i_1 \dots i_n \dots$. This is clear upon noting that $\alpha_{F_i} \in F_i([0, 1])$, and $\alpha_{F_i^k} = \alpha_{F_i}$ for each positive integer k . Hence

$$\alpha_{F_i} \in F_i^k([0, 1]) = \overbrace{f_{i_1} \dots f_{i_n} f_{i_1} \dots f_{i_n} \dots f_{i_1} \dots f_{i_n}}^{k \text{ copies of } f_{i_1} \dots f_{i_n}}([0, 1])$$

for all k . Recall that a point lies in B^+ exactly when its address is not eventually constant. Hence for every $n > 1$ and every $i \in \{0, 2\}^n$, the point α_{F_i} is a buried point. Any nonelementary subsemigroup of $\text{Aut}(\mathbb{D})$ contains many subsemigroups for which the conclusions above are valid, and so many of its attracting fixed points are also buried points in Λ^+ .

In fact our description of D^+ generalises beyond subsemigroups of $\text{Aut}(\mathbb{D})$: Suppose S is generated by $\mathcal{F} = \{f_1, \dots, f_m\}$, where each f_j lies in $\text{Aut}(\mathbb{B}^3)$. Further suppose that the limit sets of S are disjoint. Then, using ideas similar to those used above, it is possible to show that D^+ is the collection of points in Λ^+ for which there exists a sequence $i_n \in \{1, 2, \dots, m\}$ for each $n \in \mathbb{N}$, where every finite sequence taking values in $\{1, 2, \dots, m\}$ occurs somewhere along the infinite sequence $(i_1, i_2, \dots, i_n, \dots)$.

Returning to the Cantor semigroup, we show that its forward point transitive limit set, Λ_p^+ , is empty. To see this, recall that, as remarked after the proof of Theorem 4.6, for any $z \in \Lambda_p^+$ the closure of $S^{-1}(z)$ covers $\overline{\Omega^-}$. Accordingly we claim that for any $z \in \Lambda^+$ the orbit $S^{-1}(z)$ is not dense in $\overline{\Omega^-} = \overline{\mathbb{R}}$. Choose $z \in \Lambda^+$ with base-3 expansion $0.i_0 i_1 \dots$

and choose any $j = (j_0, \dots, j_{n-1}) \in \{0, 2\}^n$. If $F_j = f_{j_0} \cdots f_{j_{n-1}}$ then $F_j^{-1}(z) \in [0, 1]$ if and only if $(i_0, \dots, i_{n-1}) = (j_0, \dots, j_{n-1})$. Moreover if $(i_0, \dots, i_{n-1}) = (j_0, \dots, j_{n-1})$, then $F_j^{-1}(z) \in \Lambda^+$. Hence $S^{-1}(z)$ does not meet $[0, 1] \setminus \Lambda^+$, which proves the claim.

It should perhaps not be surprising that the Cantor semigroup has an empty point transitive limit set, as it is known (see [34, Theorem 2.3.3]) that a Kleinian group whose limit set is not the full ideal boundary necessarily has an empty point transitive limit set. We do not know if something similar is true for semidiscrete semigroups.

2.4. The limit sets of the group and inverse free parts of a semigroup. Recall from Section 1 that the inverse free part of a semigroup S is the semigroup $S \setminus S^{-1}$ of elements in S whose inverses do not belong to S . In this section we consider the limit sets of $S \setminus S^{-1}$ when S is not a group. Throughout we let G denote $S \cap S^{-1}$, the group part of S .

THEOREM 4.12. *If S is a semigroup and is not a group, then $\Lambda^+(S) = \Lambda^+(S \setminus G)$ and $\Lambda^-(S) = \Lambda^-(S \setminus G)$. Moreover we have $\Lambda_c^+(S) = \Lambda_c^+(S \setminus G)$ and $\Lambda_c^-(S) = \Lambda_c^-(S \setminus G)$.*

PROOF. Clearly $\Lambda^+(S \setminus G) \subseteq \Lambda^+(S)$. We show that $\Lambda^+(S) \subseteq \Lambda^+(S \setminus G)$. For any $z \in \Lambda^+(S)$, choose a sequence $g_n \in S$ such that

$$\lim_{n \rightarrow \infty} g_n(\mathbf{0}) = z.$$

Now choose $h \in S \setminus G$ and note that

$$\lim_{n \rightarrow \infty} g_n h(\mathbf{0}) = z.$$

Lemma 2.1 tell us that $g_n h$ lies in $S \setminus G$, so that we have $z \in \Lambda^+(S \setminus G)$. Hence we have shown that

$$\Lambda^+(S \setminus G) = \Lambda^+(S).$$

Moreover, if $g_n(\mathbf{0})$ converges to z conically, then $g_n h(\mathbf{0})$ also converges to z conically, and so this reasoning also tells us that $\Lambda_c^+(S) = \Lambda_c^+(S \setminus G)$. As usual, the corresponding statements for the backward limit sets follow upon replacing S with S^{-1} in the preceding arguments. \square

The theorem above tells us that provided a semigroup S is not a group, the limit sets and conical limit sets of the inverse free part of S are equal to the corresponding limit sets of S itself. In fact the same is true of both horospherical limit sets. In order to prove this we need a geometrical lemma. Recall that $[\alpha, \beta]$ denotes the hyperbolic geodesic between two points α and β in hyperbolic space. Similarly $(\alpha, \beta]$ denotes the hyperbolic geodesic between α and β with the point α removed. Throughout the remainder of this section, for any set $X \subseteq \mathbb{B}^3$ and $M \geq 0$ we let X_M denote the set of points that are less than a hyperbolic distance of M from some point in X .

LEMMA 4.13. *Given any $x \in \mathbb{B}^3$ and $y \in (z, x]$ we have*

$$(H_z(y))_{\rho(x,y)} = H_z(x).$$

PROOF. It is convenient to work in the upper half-space model and conjugate so that $z = \infty$. Then $H_z(y) = \{w \in \mathbb{H}^3 \mid \text{ht}[w] > \text{ht}[y]\}$ for any $y \in \mathbb{H}^3$. For any $t > 0$ the set of points that are less than a hyperbolic distance t from $H_z(y)$ is exactly the horoball $(H_z(y))_t = \{w \in \mathbb{H}^3 \mid \text{ht}[w] > \text{ht}[y - \mathbf{j}u]\}$ where $u > 0$ is such that $t = \rho(y - \mathbf{j}u, y)$. It follows that for any $y \in [x, \infty)$ we have $(H_z(y))_{\rho(x,y)} = \{w \in \mathbb{H}^3 \mid \text{ht}[w] > \text{ht}[x]\} = H_z(x)$. \square

THEOREM 4.14. *If S is a semigroup and is not a group, then*

$$\Lambda_h^+(S \setminus G) = \Lambda_h^+(S).$$

PROOF. Let $\Lambda_{h,x}^+(S)$ denote the points $z \in \Lambda^+(S)$ such that some horoball based at z does not meet the S -orbit of x . It is easy to see that

$$\Lambda_{h,x}^+(S) = \Lambda_h^+(S)$$

for all $x \in \mathbb{B}^3$. Using Lemma 4.13 we show that

$$\Lambda_{h,x}^+(S \setminus G) = \Lambda_{h,x}^+(S)$$

for any $x \in \mathbb{B}^3$.

Given any $x \in \mathbb{B}^3$ we certainly have

$$\Lambda_{h,x}^+(S \setminus G) \subseteq \Lambda_{h,x}^+(S).$$

To show the reverse inclusion suppose z is a point in $\Lambda_{h,x}^+(S)$ and g_n is a sequence in S such that the sequence of points $g_n(x)$ meets any horoball based at z . By assumption we can choose some $f \notin G$, and again appealing to Lemma 2.1, we see that the element g_nf lies in $S \setminus G$, itself a semigroup. Moreover

$$\rho(g_nf(x), g_n(x)) = \rho(f(x), x).$$

Let $H_z(u)$ be any horoball based at z . Then by the claim we can choose $v \in \mathbb{B}^3$ such that

$$(H_z(v))_{\rho(f(0),0)} \subseteq H_z(u).$$

Since $g_n(x)$ meets $H_z(v)$ it follows $g_nf(x)$ meets $H_z(u)$. Since $H_z(u)$ was arbitrary we have a sequence of points $g_nf(x)$ in the $S \setminus G$ orbit of x which meets every horoball based at z . Hence we have shown that

$$\Lambda_{h,x}^+(S \setminus G) = \Lambda_{h,x}^+(S),$$

and so

$$\Lambda_h^+(S \setminus G) = \Lambda_h^+(S).$$

□

THEOREM 4.15. *Either $\Lambda^+(S) = \Lambda(G)$ or $\Lambda(G) \subseteq \overline{\Lambda^+ \setminus \Lambda(G)}$. In other words $\Lambda^+(S) = \Lambda(G)$ or $\Lambda^+ = \overline{\Lambda^+ \setminus \Lambda(G)}$.*

PROOF. Each $g \in G$ fixes Λ setwise, and so each g maps $\Lambda^+(S) \setminus \Lambda(G)$, and hence its closure, into itself. Now either $\overline{\Lambda^+(S) \setminus \Lambda(G)}$ has less than two points, in which case $\Lambda^+(S) = \Lambda(G)$, or as the smallest G -forward invariant set containing at least two points, $\Lambda^+(G)$ is contained in $\overline{\Lambda^+(S) \setminus \Lambda(G)}$. But $\Lambda^+(G) = \Lambda(G)$ which gives the result. □

As usual the analogous version of the theorem above for the backward limit set follows upon replacing S with S^{-1} . Shown in Figure 4.3 are the limit sets of the semidiscrete semigroup found by taking the Kleinian group G generated by (see [33, page 272]) the two transformations

$$z \mapsto \frac{(0.8210 - 0.3832i)z + (0.3832 + 1.8210i)}{(-0.3832 + 0.1789i)z + (0.8210 - 0.3832)}$$

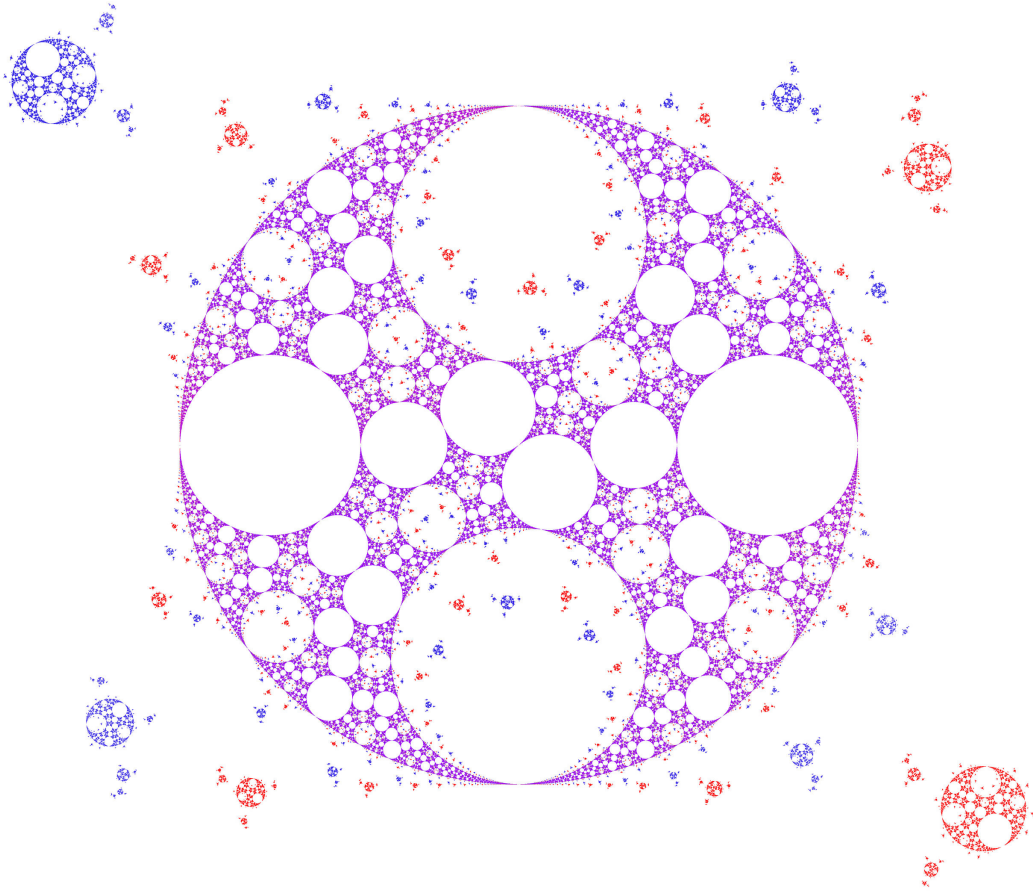


FIGURE 4.3. Limit sets of a semigroup with nonempty group and inverse free parts

and

$$z \mapsto \frac{(1-i)z + 1}{z + (1+i)},$$

then, in the spirit of Theorem 2.20, appending the generator

$$z \mapsto \frac{(2.77 - 1.63i)z + (2.25 - 5.22i)}{(1.74 + 0.0213i)z + (2.77 - 1.63i)}.$$

The forward limit set is in red, while the backward limit set is in blue. They intersect on the purple set, which is the limit set of the Kleinian group G . Each point in the purple set is a limit point of the red set, and a limit point of the blue set.

In general, if a semidiscrete semigroup S has a nonelementary group part G , then $\Lambda^+(S) \cap \Lambda^-(S)$ contains $\Lambda(G)$, necessarily a perfect set. It is an open problem, even in two dimensions, whether or not the converse is true. That is, if S is a semidiscrete semigroup and $\Lambda^+(S) \cap \Lambda^-(S)$ is a perfect set, must S necessarily have a nonelementary group part? This question is particularly interesting in view of Theorems 3.33 and 3.35.

3. On semigroups with disjoint limit sets

In this section we study semidiscrete and inverse free semigroups that are finitely-generated and have disjoint limit sets. We shall show in Theorem 4.16 that if a finitely-generated semigroup has disjoint limit sets, then, provided it has no elliptic generators, it is semidiscrete and inverse free too. Semigroups that belong to this class are very well behaved, even in higher dimensions. For example, in [15] the authors show that given a finitely-generated semigroup S acting in three dimensions and with disjoint limit sets, there exists a metric (which they construct) defined on some open set containing Λ^+ , with respect to which the semigroup can be regarded as a contracting iterated function system with limit set Λ^+ . We shall show that if we further suppose S is semidiscrete and inverse free, then there exists a finitely-generated subsemigroup of S that can be regarded as a contracting iterated function system with respect to the chordal metric on some open set containing Λ^+ , and which has limit set Λ^+ .

In Chapter 3 we characterised semidiscrete and inverse free semigroups in two dimensions. In higher dimensions we have the following.

THEOREM 4.16. *Suppose \mathcal{F} is a finite subset of $\text{Aut}(\mathbb{B}^3)$ such that the identity and any elliptic maps in \mathcal{F} generate a group of finite order. If the limit sets of $S = \langle \mathcal{F} \rangle$ are disjoint, then S is semidiscrete. If moreover \mathcal{F} does not contain the identity or any elliptic maps, then S is inverse free.*

PROOF. Since the limit sets of S are disjoint, \mathcal{F} contains no parabolic elements. The identity and any elliptic elements in \mathcal{F} fix Λ^+ setwise, as their inverses also lie in S and Λ^+ is forward invariant. Any other elements in \mathcal{F} are loxodromic. If any loxodromic $f \in \mathcal{F}$ were to fix Λ^+ , then both fixed points of f must lie in Λ^+ . This is not possible as

the repelling fixed point of f lies in Λ^- , which by assumption does not meet Λ^+ . Hence each loxodromic element in \mathcal{F} maps Λ^+ strictly inside itself. It now follows that S is semidiscrete by Theorem 2.5. If moreover \mathcal{F} does not contain the identity or any elliptic maps, then \mathcal{F} must be purely loxodromic and no element in \mathcal{F} fixes Λ^+ setwise. Hence, again by Theorem 2.5, S is also inverse free. \square

For any $f \in \text{Aut}(\mathbb{B}^3)$ we define the quantity $L(f)$ by

$$L(f) = \sup_{\substack{z, w \in \mathbb{S}^2 \\ z \neq w}} \frac{\chi(f(z), f(w))}{\chi(z, w)},$$

and shall make use of the following equation given in [3, Theorem 3.6.1]

$$(12) \quad L(f) = \exp[\rho(\mathbf{0}, f(\mathbf{0}))].$$

We shall make repeated use of the following important property enjoyed by all semigroups generated by a bounded set. Recall that $\overline{\mathcal{F}}$ denotes the closure of \mathcal{F} in $\text{Aut}(\mathbb{B}^3)$.

LEMMA 4.17. *Let S be a semigroup generated by some bounded set \mathcal{F} . Then*

$$\Lambda^+(S) = \bigcup_{f \in \overline{\mathcal{F}}} f(\Lambda^+(S)).$$

PROOF. One inclusion is clear from the forward invariance of $\Lambda^+ = \Lambda^+(S)$. For the other inclusion, suppose $z \in \Lambda^+$ and fix some sequence $g_n \in S$ such that $g_n(\mathbf{0}) \rightarrow z$. We show z lies in $f(\Lambda^+)$ for some $f \in \overline{\mathcal{F}}$. For each n we can choose $f_n \in \mathcal{F}$ and $h_n \in S \cup \{I\}$ such that $g_n = f_n h_n$. By compactness we can pass to a subsequence, without loss of generality the sequence itself, such that $f_n \rightarrow f \in \overline{\mathcal{F}}$. Now it follows that $\chi(f_n h_n(\mathbf{0}), z) \rightarrow 0$ as $n \rightarrow \infty$, and so

$$\frac{1}{L(f_n^{-1})} \chi(h_n(\mathbf{0}), f_n^{-1}(z)) \rightarrow 0 \quad \text{as } n \rightarrow \infty,$$

by the definition of $L(f_n^{-1})$. Making use of equation (12), for each $f \in \text{Aut}(\mathbb{B}^3)$ we have $L(f) = \exp[\rho(\mathbf{0}, f(\mathbf{0}))] = \exp[\rho(f^{-1}(\mathbf{0}), \mathbf{0})] = L(f^{-1})$, and as \mathcal{F} is bounded, it follows that $\{L(f) \mid f \in \mathcal{F}\}$ is bounded above. Hence $\chi(h_n(\mathbf{0}), f^{-1}(z)) \rightarrow 0$ as $n \rightarrow \infty$, and so $f^{-1}(z) \in \Lambda^+$. We have shown that for all $z \in \Lambda^+$ there exists $f \in \overline{\mathcal{F}}$ such that $z \in f(\Lambda^+)$. Hence we have proven the other inclusion. \square

THEOREM 4.18. *Let S be a semigroup generated by a finite set \mathcal{F} . If S is semidiscrete and inverse free and the limit sets of S are disjoint, then for each point p in Λ^+ there is a composition sequence generated by \mathcal{F} that converges conically to p . Moreover, every composition sequence converges conically to a single point on the ideal boundary, and in particular, converges ideally.*

PROOF. We first show that each point z in Λ^+ also lies in Λ_c^+ . By Lemma 4.17 there exists $f_1 \in \mathcal{F}$ and $z_1 \in \Lambda^+$ such that $z = z_0 = f_1(z_1)$. Similarly there exists $f_2 \in \mathcal{F}$ and $z_2 \in \Lambda^+$ such that $z_1 = f_2(z_2)$ so that $z = f_1 f_2(z_2)$. Continuing in this way we recursively choose $f_n \in \mathcal{F}$ and $z_n \in \Lambda^+$ such that $z_{n-1} = f_n(z_n)$. Writing $F_n = f_1 \cdots f_n$ we have $z = F_n(z_n)$ so that

$$z_n = F_n^{-1}(z).$$

Since $I \notin \bar{S}$ the sequence F_n is escaping. By Theorem 1.2 if z did not lie in the backward conical limit set of the sequence F_n^{-1} , then z_n must accumulate only in Λ^- ; but this cannot be as z_n is a sequence in Λ^+ . Hence we must have

$$z \in \Lambda_c^-(F_n^{-1}) = \Lambda_c^+(F_n) \subseteq \Lambda_c^+(S)$$

as claimed. In fact F_n converges to z ideally. Indeed, we show that every composition sequence converges ideally, as the theorem also asserts. To see this, consider $\mathcal{C}(\Lambda^+)$, the convex hull of Λ^+ , that is the smallest hyperbolically convex subset of $\overline{\mathbb{B}^3}$ containing every geodesic whose landing points both belongs to Λ^+ . As Λ^+ is forward invariant under S , so is $\mathcal{C}(\Lambda^+)$. Since the limit sets are disjoint and F_n^{-1} is escaping, $\rho(F_n^{-1}(\mathbf{0}), \mathcal{C}(\Lambda^+)) \rightarrow \infty$ as $n \rightarrow \infty$. Hence $\rho(\mathbf{0}, F_n(\mathcal{C}(\Lambda^+))) = \rho(\mathbf{0}, \mathcal{C}(F_n(\Lambda^+))) \rightarrow \infty$ as $n \rightarrow \infty$. Since $F_n(\Lambda^+)$ is a decreasing nested sequence of sets and its chordal diameter converges to 0, z is the only accumulation point of F_n . It follows that $F_n(\mathbf{0})$ converges conically to z . \square

Recall that for a semigroup S , the set Λ_q^+ is the collection of points in Λ^+ that are limit points of some composition sequence generated by S . Hence for a semigroup S as above, $\Lambda^+ = \Lambda_c^+ = \Lambda_q^+$. Furthermore, since S has disjoint limit sets and is semidiscrete and inverse free, then S^{-1} also enjoys these properties. It follows that we also have $\Lambda^- = \Lambda_c^- = \Lambda_q^-$. We cannot drop the assumption that \mathcal{F} is bounded, as the following example demonstrates.

Take loxodromic elements g and h in $\text{Aut}(\mathbb{D})$ whose axes are configured as in Figure 4.4 below.

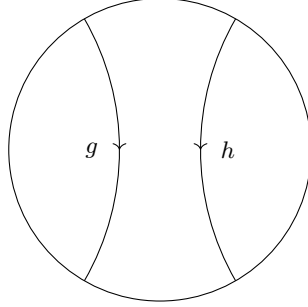


FIGURE 4.4. Loxodromic maps g and h

Let X be the complementary component of $\mathbb{S}^1 \setminus \{\alpha_g, \alpha_h\}$ that does not meet either generator's repelling fixed point. Let \mathcal{F} be the unbounded set $\{gh, g^2h^2, g^3h^3, \dots\}$ and let S be the semigroup generated by \mathcal{F} . Then $\alpha_g \in \Lambda^+(S)$ since $g^n h^n(0) \rightarrow \alpha_g$ as $n \rightarrow \infty$. Yet for any composition sequence F_n generated by \mathcal{F} , we have, for all n , $F_n(X) = f_1 \cdots f_n(X) \subseteq f_1(X) = g^m h^m(X)$ for some $m \in \mathbb{N}$. Hence $F_n(0)$ can only accumulate in $g^m h^m(X)$, and in particular, $F_n(0)$ does not accumulate at α_g . This means that $\Lambda_q^+(S)$ is not equal to $\Lambda^+(S)$ for this particular example.

3.1. A sufficient condition for discreteness. In [10, Theorem 10] the author shows that if a finitely-generated subsemigroup of $\text{Aut}(\mathbb{D})$ has disjoint limit sets, then the semigroup must be discrete. In this subsection we show that the same is true in all dimensions if we additionally suppose S is generated by a set of loxodromic elements. Recall that if X is a subset of a topological space, its derived set, denoted X' , is the set of limit points of X . Hence a semigroup S is discrete precisely when S' is empty. It is easy to check that \overline{S} is also a semigroup whose limit sets coincide with those of S . Motivated by this, we begin by studying S' , which it turns out is a semigroup in its own right. For any $A, B \subseteq \text{Aut}(\mathbb{B}^3)$ we define AB to be the set $\{ab \mid a \in A \text{ and } b \in B\}$.

THEOREM 4.19. *Suppose S is a semigroup, then*

- (i) S' is a semigroup,
- (ii) SS' and $S'S$ are subsets of S' , and

(iii) if S' is nonempty then $\Lambda^+(S) = \Lambda^+(S')$ and $\Lambda_c^+(S) = \Lambda_c^+(S')$. Similarly $\Lambda^-(S) = \Lambda^-(S')$ and $\Lambda_c^-(S) = \Lambda_c^-(S')$.

PROOF. We must show S' is closed under composition. Suppose that $g, h \in S'$, say

$$g = \lim_{n \rightarrow \infty} g_n$$

and

$$h = \lim_{n \rightarrow \infty} h_n$$

where g_n and h_n are sequences of distinct elements in $S \setminus \{g\}$ and $S \setminus \{h\}$ respectively. Since the group of Möbius transformations is a topological group we have

$$gh = \lim_{n \rightarrow \infty} g_n h_n.$$

If $g_n h_n$ contains a subsequence of distinct elements in $S \setminus \{gh\}$ then $gh \in S'$. Otherwise $g_n h_n = gh$ for all large enough n . Since $h_n \neq h_{n+1}$ we must have $g_n h_{n+1} \neq gh$, and as

$$gh = \lim_{n \rightarrow \infty} g_n h_{n+1}$$

we have $gh \in S'$ in this case too.

To see $SS' \subseteq S'$ suppose again that $g \in S'$ via

$$g = \lim_{n \rightarrow \infty} g_n$$

where g_n are distinct elements in $S \setminus \{g\}$. Now for any $h \in S$, hg_n is a sequence of distinct points in $S \setminus \{hg\}$ and converges to hg . Hence hg lies in S' as claimed.

Now suppose S has a nonempty derived set. We first show that $\Lambda^+(S) \subseteq \Lambda^+(S')$. Working in the ball model, suppose $z \in \Lambda^+(S)$ and choose a sequence f_n in S such that

$$\lim_{n \rightarrow \infty} f_n(\mathbf{0}) = z.$$

Since S' is nonempty we can choose some $g \in S'$. Since

$$\lim_{n \rightarrow \infty} f_n g(\mathbf{0}) = z$$

and $SS' \subseteq S'$, it follows that $z \in \Lambda^+(S')$.

For the converse, suppose z lies in $\Lambda^+(S')$ and h_n is a sequence in S' such that

$$\lim_{n \rightarrow \infty} h_n(\mathbf{0}) = z.$$

Since each h_n lies in S' we can choose a sequence g_n in S such that $\sigma(g_n, h_n) < 1/n$ for each $n \in \mathbb{N}$. Hence $\chi(g_n(\mathbf{0}), h_n(\mathbf{0})) < 1/n$, so that $g_n(\mathbf{0})$ also converges to z , and we see that $z \in \Lambda^+$.

To show the conical limit sets of S and S' coincide, we suppose that f_n is a sequence in S and that $f_n(\mathbf{0}) \rightarrow z$ conically as $n \rightarrow \infty$. Since $\rho(f_n g(\mathbf{0}), f_n(\mathbf{0})) = \rho(g(\mathbf{0}), \mathbf{0})$ it follows that for any $g \in S'$ $f_n g(\mathbf{0})$ converges to z conically. Since SS' is contained in S' we see that $\Lambda_c^+(S) \subseteq \Lambda_c^+(S')$. For the reverse inclusion we use a version of the argument we used to show $\Lambda^+(S) \subseteq \Lambda^+(S')$: Suppose z lies in $\Lambda_c^+(S')$ via some sequence h_n in S' such that $h_n(\mathbf{0})$ converges to z conically. Since the map that sends any $g \in \text{Aut}(\mathbb{B}^3)$ to $g(\mathbf{0}) \in \mathbb{B}^3$ is continuous, it follows that we can choose $g_n \in S$ such that $\rho(g_n(\mathbf{0}), h_n(\mathbf{0})) < 1$. Hence $g_n(\mathbf{0})$ converges to z conically, and so z lies in $\Lambda_c^+(S)$.

As usual the results for the backward limit sets follow upon considering S^{-1} instead of S , and noting that the map $g \rightarrow g^{-1}$ from $\text{Aut}(\mathbb{B}^3)$ to itself is a homeomorphism. \square

Before proving the main result of this subsection, we need a preparatory lemma.

LEMMA 4.20. *Suppose S is a semigroup generated by a finite set \mathcal{F} . Then for all $g \in S'$ there exists $f \in \mathcal{F}$ such that $f^{-1}g \in S'$.*

PROOF. Given any $g \in S'$ we can choose a sequence of distinct elements $g_n \in S \setminus \{g\}$, such that $g_n \rightarrow g$. Since \mathcal{F} is finite, for some $f \in \mathcal{F}$ we can pass to a subsequence, which we do not rename, such that

$$g_n = fh_n,$$

where h_n is a sequence of distinct elements in $S \setminus \{f^{-1}g\}$. Since $h_n \rightarrow f^{-1}g$ it follows that $f^{-1}g \in S'$ as claimed. \square

THEOREM 4.21. *Let S be a semigroup generated by a finite set \mathcal{F} . If $I \notin \overline{S}$ and the limit sets Λ^+ and Λ^- are disjoint, then S is discrete.*

PROOF. We show that if S is not discrete then Λ^+ and Λ^- must meet. If S is not discrete then S' is nonempty so we can choose $g \in S'$. Repeatedly applying Lemma 4.20 above, we can generate a sequence $f_n \in \mathcal{F}$ such that

$$f_n^{-1} \cdots f_1^{-1} g \in S'$$

for each n . We let $F_n = f_1 \cdots f_n$ for each $n \in \mathbb{N}$. Since $I \notin \overline{S}$, the composition sequence F_n is escaping by Theorem 2.7; hence both $F_n^{-1}g$ and F_n^{-1} are escaping sequences, and in particular have at least one accumulation point on \mathbb{S}^2 . The sequence $F_n^{-1}g$ lies in S' while F_n^{-1} lies in S^{-1} . Since

$$\rho(F_n^{-1}g(\mathbf{0}), F_n^{-1}(\mathbf{0})) = \rho(g(\mathbf{0}), (\mathbf{0}))$$

any accumulation point ζ of $F_n^{-1}(\mathbf{0})$ is an accumulation point of $F_n^{-1}g(\mathbf{0})$, so that ζ lies in both $\Lambda^-(S)$ and $\Lambda^+(S')$. Finally by Proposition 4.19 $\Lambda^+(S') = \Lambda^+(S)$, and so $\Lambda^+(S)$ meets $\Lambda^-(S)$ at ζ . \square

3.2. Perturbed semigroups. In the following subsection we investigate what happens to the limit sets of a semigroup $S = \langle \mathcal{F} \rangle$ when its generating set \mathcal{F} , considered as an $|\mathcal{F}|$ -tuple of Möbius transformations, is perturbed. By perturbing a semigroup we mean perturbing each of its generators in $\text{Aut}(\mathbb{B}^3)$. Most of the results given here assume that \mathcal{F} is finite, although we occasionally only require \mathcal{F} to be bounded.

Suppose that S is a finitely-generated, semidiscrete and inverse free semigroup. Our first theorem indicates that if the limit sets of S are disjoint, then they behave nicely as the generators of S are perturbed. Before stating this theorem more precisely, we first give an important lemma, which we shall use again at the close of this chapter.

LEMMA 4.22. *Let S be a semidiscrete and inverse free semigroup generated by a finite set \mathcal{F} . Then for all $M > 0$ there exists a finite subset \mathcal{F}_M of S , such that for all $f \in \mathcal{F}_M$, $\rho(f(\mathbf{0}), \mathbf{0}) > M$, and $\Lambda^+(S) = \Lambda^+(S_M)$ where S_M is the semigroup generated by \mathcal{F}_M . In*

particular,

$$\Lambda^+(S) = \bigcup_{f \in \mathcal{F}_M} f(\Lambda^+(S)).$$

PROOF. Choose $M > 0$ and let us write $N = \{0, 1, \dots, N-1\}$, so that N^n is the set of n tuples on the set $\{0, 1, \dots, N-1\}$. Let $T = N^{<\omega}$ be the union $\bigcup_{n=1}^{\infty} N^n$, which we endow with the usual partial order relation, that is, for $s, t \in T$, then $s \leq t$ if and only if the domain of s is contained in the domain of t , and s agrees with t on the domain of s . Equipped with this partial ordering T is a tree. Letting $\mathcal{F} = \{f_0, \dots, f_{N-1}\}$, the elements of T may be regarded as words in \mathcal{F} , and so as elements of S . More precisely we make use of the projection

$$\pi : T \rightarrow S$$

given by

$$\pi((i_0, \dots, i_{N-1})) = f_{i_0} \cdots f_{i_{N-1}}.$$

The projection is injective exactly when S is free. The n^{th} level of T is the set N^n , which corresponds to the collection of n -length words generated from \mathcal{F} . The level to which an element in T belongs is called its *height*, and we write $\text{ht}(t)$ to denote the height of $t \in T$. We say two elements $s, t \in T$ are *comparable* if $s \leq t$ or $t \leq s$, otherwise they are *incomparable*. A *cofinal branch* of T is any maximal, linearly ordered subset of T that meets every level.

To prove the lemma we find a set $T_M \subseteq T$ such that;

- (i) T_M is a finite set of pairwise incomparable elements,
- (ii) $\rho(\pi(t)(\mathbf{0}), \mathbf{0}) > M$ for each $t \in T_M$,
- (iii) for each $t \in T$ there is some (unique by (i), although we do not use uniqueness) $s \in T_M$ such that s and t are comparable, and,
- (iv)

$$\Lambda^+ = \bigcup_{t \in T_M} \pi(t)(\Lambda^+).$$

In order to find T_M , we recursively construct a monotonically increasing nest of sets $T_M(n) \subseteq T$. We define $T_M(0) = \emptyset$. Suppose that for some $n \in \mathbb{N}$ the set $T_M(n-1)$ has

been constructed. Now consider all elements in the n^{th} level of T which are not greater than some element in $T_M(n-1)$, that is the set

$$L_n = N^n \setminus \{t \in T \mid s \leq t \text{ for some } s \in T_M(n-1)\}.$$

Now put

$$T_M(n) = T_M(n-1) \cup \{t \in L_n \mid \rho(\pi(t)(\mathbf{0}), \mathbf{0}) > M\}.$$

We show that L_n is empty for some n . For if not, then

$$T' = \bigcup_{n \in \mathbb{N}} L_n$$

is a tree of height ω . We claim T' is a tree. To see this, it is enough to show that for any $t \in T'$ and $s \leq t$ we have $s \in T'$. Suppose towards contradiction that $s \notin T'$. Then $s \notin L_{\text{ht}(s)}$, so that $r \leq s$ for some $r \in T_M(\text{ht}(s)-1) \subseteq T_M(\text{ht}(t)-1)$. But then $t \leq s \leq r$ which contradicts $t \in L_{\text{ht}(t)}$. Hence T' is a tree as claimed. Since T' has height ω and each level is finite, König's lemma (see [23, Lemma 14.2]) implies it contains a linearly ordered subset of T that intersects every level, $j \in N^\omega$ say. The associated sequence of Möbius transformations has n^{th} term $F_n = f_{j_0} \cdots f_{j_{n-1}}$. This is a composition sequence satisfying $\rho(F_n(\mathbf{0}), \mathbf{0}) \leq M$ for all n , contradicting our assumption that S is semidiscrete and inverse free by Theorem 2.7. Hence each L_n is eventually empty after all. This means that there is a set, which we take to be T_M , such that $T_M(n) = T_M$ for all large enough n . We set $\mathcal{F}_M = \pi(T_M)$, and verify that T_M has all four properties claimed. To see (i), since each L_n is eventually empty, T_M is finite. For each n when constructing the set $T_M(n)$ we only append elements to $T_M(n-1)$ that are incomparable with each other, and with every element already in $T_M(n-1)$. It follows that any two elements in T_M are incomparable, giving (i). Property (ii) clearly holds by the construction. If (iii) were false, then for some $t \in T$ there would be no $s \in T_M$ such that $s < t$. But then L_n is not empty for each $n \geq \text{ht}(t)$, a contradiction. To see (iv), first let h be the maximum of the heights of elements in T_M , and choose any $z \in \Lambda^+$. Then there is a sequence $g_n \in S$ such that $g_n(\mathbf{0}) \rightarrow z$ as $n \rightarrow \infty$. Now by property (iii) any $t \in T$ may be written as the concatenation of elements in T_M , followed by some element of height no more than h .

Since each Möbius map g_n lifts to an element in T , each g_n can be expressed as

$$g_n = f_{i(1,n)} f_{i(2,n)} \cdots f_{i(l(n),n)} r_n,$$

where each $f_{i(j,n)}$ belongs to T_M , $i(m,n) \in N$ and $r_n \in S$ where r_n , regarded as a word in \mathcal{F} , has length no more than h . Since \mathcal{F} is finite, the set of compositions of no more than h elements from \mathcal{F} forms a bounded set, hence

$$\lim_{n \rightarrow \infty} f_{i(1,n)} f_{i(2,n)} \cdots f_{i(l(n),n)}(\mathbf{0}) = z.$$

This shows that $\Lambda^+(S) \subseteq \Lambda^+(S_M)$, and as the reverse inclusion is clear we have $\Lambda^+(S) = \Lambda^+(S_M)$. Since \mathcal{F}_M is finite we can apply Lemma 4.17 to the semigroup S_M and infer (iv). \square

This lemma verifies the statement made in the introduction to this section, that a finitely-generated, semidiscrete and inverse free semigroup with disjoint limits sets can be regarded as a contracting iterated function system with respect to the chordal metric on some open neighbourhood of Λ^+ , and with limit set Λ^+ . To see this, first recall from Chapter 1 that for a loxodromic map f , the *isometric disc* of f , which we shall denote by $I^+(f)$, is the open chordal disc contained in $\overline{\mathbb{C}}$ where the chordal derivative of f is greater than 1. We let $I^-(f)$ denote the isometric disc of f^{-1} , and note that $I^-(f)$, contains α_f while $I^+(f)$ contains β_f . Both isometric discs have radius $1/\sinh[\frac{1}{2}\rho(f(\mathbf{0}), \mathbf{0})]$. The action of f is to map the interior of the complement of $I^+(f)$ onto $I^-(f)$. If $S = \langle \mathcal{F} \rangle$ is a finitely-generated, semidiscrete and inverse free semigroup with disjoint limits sets, then we can choose M large enough such that the isometric discs of each map in \mathcal{F}_M do not meet Λ^+ . Then $S_M = \langle \mathcal{F}_M \rangle$ is a subsemigroup of S and inherits its limit sets. Moreover each $f \in \mathcal{F}_M$ is a strict contraction (with respect to the chordal metric) on Λ^+ . This shows that given a finitely-generated semidiscrete and inverse free semigroup whose limit sets are disjoint and some positive real number M , we can always find a subsemigroup whose generators all have norm at least M and whose limit sets are equal to those of the original semigroup.

So far in this thesis, we have regarded the generating set of a semigroup as an unordered set of Möbius transformations. In this subsection it is sometimes convenient to enumerate \mathcal{F} as a (finite) sequence, and instead regard the generators of a semigroup as a point in

$\text{Aut}(\mathbb{B}^3)^{|\mathcal{F}|}$. To emphasise this new point of view, instead of referring to $\mathcal{F} \subseteq \text{Aut}(\mathbb{B}^3)$ as a generating set, we shall say \mathcal{F} is a *generating sequence*. We shall say that the generating sequence \mathcal{F} is bounded if the range of \mathcal{F} is bounded in $\text{Aut}(\mathbb{B}^3)$, and we shall denote by $\langle \mathcal{F} \rangle$ the semigroup generated by the range of \mathcal{F} . For $k \in \mathbb{N}$ we endow $\text{Aut}(\mathbb{B}^3)^k$ with a product metric: we define the distance between points (f_1, \dots, f_k) and (g_1, \dots, g_k) in $\text{Aut}(\mathbb{B}^3)^k$ by $\max_{i=1, \dots, k} \sigma(f_i, g_i)$. In fact the choice of metric is not particularly important and we shall not need its explicit form again. Indeed, any of the usual product metrics which induce the same topology as the one above will suffice for our purposes.

THEOREM 4.23. *Suppose S is generated by the finite set $\{f_1, \dots, f_k\} \subseteq \text{Aut}(\mathbb{B}^3)$ and that S is semidiscrete and inverse free. If the limit sets of S are disjoint, then in some neighbourhood of $\mathcal{F} = (f_1, \dots, f_k) \in \text{Aut}(\mathbb{B}^3)^k$ (with respect to the product topology on $\text{Aut}(\mathbb{B}^3)$) the functions from $\text{Aut}(\mathbb{B}^3)^k$ to the collection of compact subsets of \mathbb{S}^2 (endowed with the usual Hausdorff metric) given by*

$$\mathcal{F} \mapsto \Lambda^+(\langle \mathcal{F} \rangle) \text{ and } \mathcal{F} \mapsto \Lambda^-(\langle \mathcal{F} \rangle)$$

are continuous.

PROOF. By symmetry of argument we only need to show that the function $\mathcal{F} \mapsto \Lambda^+(\langle \mathcal{F} \rangle)$ is continuous. We show that for all small enough $\epsilon > 0$ there exists $\delta > 0$ such that whenever \mathcal{G} is within δ of \mathcal{F} in $\text{Aut}(\mathbb{B}^3)^k$, we have $\Lambda^+(\langle \mathcal{G} \rangle)$ within ϵ of $\Lambda^+(\langle \mathcal{F} \rangle)$ in the Hausdorff metric. Let $\pi_{\mathcal{F}} : k^{<\omega} \rightarrow \langle \mathcal{F} \rangle$ denote the projection map that sends an element in the tree $k^{<\omega}$ to its associated Möbius transformation in $\langle \mathcal{F} \rangle$. By Lemma 4.22 there exists $T \subseteq k^{<\omega}$ such that $\Lambda^+(\langle \pi_{\mathcal{F}}(T) \rangle) = \Lambda^+(\langle \mathcal{F} \rangle)$ and such that for all $t \in T$, the isometric disc of $\pi_{\mathcal{F}}(t)$ has diameter smaller than $\epsilon/6$. We shall assume that the distance between $\Lambda^+(\langle \mathcal{F} \rangle)$ and $\Lambda^-(\langle \mathcal{F} \rangle)$ is larger than 3ϵ , so that for no $s, t \in T$ does $I^-(\pi_{\mathcal{F}}(s))$ meet $I^+(\pi_{\mathcal{F}}(t))$. As f is contracting with respect to the chordal metric in the interior of $\mathbb{S}^2 \setminus I^+(f)$, the set \mathcal{F} is a contracting iterated function system on

$$\mathbb{S}^2 \setminus \bigcup_{f \in \mathcal{F}} I^+(f)$$

(with respect to the chordal metric) and has limit set $\Lambda^+(\langle \mathcal{F} \rangle)$, which is contained in

$$\bigcup_{f \in \mathcal{F}} I^-(f).$$

Each point in $\Lambda^+(\langle \mathcal{F} \rangle)$ is certainly within $\epsilon/6$ of some point in $\bigcup_{f \in \mathcal{F}} I^-(f)$. Conversely, given any $f \in \mathcal{F}$ and any point $z \in I^-(f)$, the point α_f , which lies in $\Lambda^+(\langle \mathcal{F} \rangle)$, is within $\epsilon/6$ of z . Hence z lies in the $\epsilon/6$ neighbourhood of $\Lambda^+(\langle \mathcal{F} \rangle)$ and it follows that

$$(13) \quad d_H \left(\bigcup_{t \in T} I^-(\pi_{\mathcal{F}}(t)), \Lambda^+(\langle \mathcal{F} \rangle) \right) < \epsilon/6.$$

The choice of T ensures that for any $\mathcal{G} \in \text{Aut}(\mathbb{B}^3)^k$ we have $\Lambda^+(\langle \pi_{\mathcal{G}}(T) \rangle) = \Lambda^+(\langle \mathcal{G} \rangle)$. We choose an open neighbourhood U of \mathcal{F} small enough such that for each $\mathcal{G} \in U$ and $t \in T$ the disc $I^-(\pi_{\mathcal{G}}(t))$ is within $\epsilon/3$ of $I^-(\pi_{\mathcal{F}}(t))$ and $I^+(\pi_{\mathcal{G}}(t))$ is within $\epsilon/3$ of $I^+(\pi_{\mathcal{F}}(t))$ (both with respect to the Hausdorff metric).

Hence we have

$$(14) \quad d_H \left(\bigcup_{t \in T} I^-(\pi_{\mathcal{F}}(t)), \bigcup_{t \in T} I^-(\pi_{\mathcal{G}}(t)) \right) < \epsilon/3.$$

Moreover, we can ensure U is small enough such that the diameter of each $I^-(\pi_{\mathcal{G}}(t))$ is less than $\epsilon/3$, and so by the same argument used to derive (13) we see that

$$(15) \quad d_H \left(\bigcup_{t \in T} I^-(\pi_{\mathcal{G}}(t)), \Lambda^+(\langle \mathcal{G} \rangle) \right) < \epsilon/3.$$

Hence by the triangle inequality the relations (13), (14) and (15) imply that

$$d_H \left(\Lambda^+(\langle \mathcal{F} \rangle), \Lambda^+(\langle \mathcal{G} \rangle) \right) < \epsilon.$$

This shows that $\mathcal{F} \mapsto \Lambda^+(\langle \mathcal{F} \rangle)$ is continuous. By symmetry of argument $\mathcal{F} \mapsto \Lambda^-(\langle \mathcal{F} \rangle)$ is continuous as well. \square

The theorem can fail if the limit sets are not disjoint. For example, consider the subsemi-group of $\text{Aut}(\mathbb{D})$ generated by two antiparallel loxodromic transformations g and h defined as follows. Let σ be a reflection in the geodesic $\ell = [-i, i]$, and let σ_g and σ_h be reflections in the geodesics $\ell_g = [z, 1]$ and $\ell_h = [\bar{z}, 1]$ respectively, where $z = e^{i\theta}$ for some $\theta \in (0, \pi/2)$.

Let $g = \sigma_g \sigma$ and $h = \sigma \sigma_h$ so that $gh = \sigma_g \sigma_h$ is a parabolic map that fixes 1. Hence the limit sets of the semigroup $S = \langle g, h \rangle$ meet at 1. However in every neighbourhood of g there exists a map g_1 such that $g_1 h$ is elliptic of infinite order. Since g_1 and h do not commute, $S_1 = \langle g_1, h \rangle$ is dense in $\text{Aut}(\mathbb{D})$ by [2], and so both limit sets of S_1 (and in particular $\Lambda^+(S_1)$) equal the whole ideal boundary. However $\Lambda^+(S)$ is a proper subset of the ideal boundary. To see this, let J_g be the closed arc contained in the unit circle whose end points are the landing points of ℓ_g and which contains α_g . Similarly let J_h be the arc contained in the unit circle whose end points are the landing points of $\sigma(\ell_h)$ and which contains α_h . It can be seen that both g and h map the set $J_g \cup J_h$ (the two highlighted arcs in Figure 4.5) within itself, and so $\Lambda^+(S) \subseteq J_g \cup J_h$.

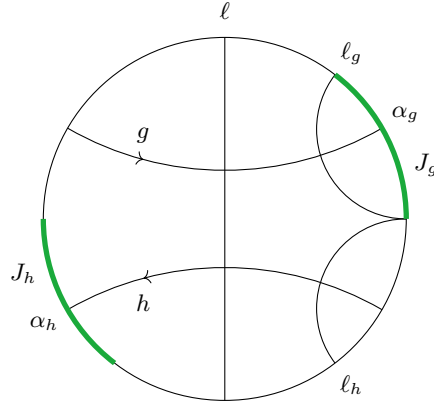


FIGURE 4.5. The generators g, h and the arcs J_g and J_h

Statements analogous to Theorem 4.23 can be found within the theory of complex dynamics. Let $\text{Rat}(d)$ be the space of rational maps of degree d , endowed with the compact open topology. A rational map $f \in \text{Rat}(d)$ is *hyperbolic* if there exists a smooth conformal metric defined on an open set containing the Julia set $J(f)$, upon which f is expanding. It can be shown that for a hyperbolic map f , there is some neighbourhood of f in $\text{Rat}(d)$ (endowed with the compact-open topology) upon which $J(f)$ moves continuously with f . See Branner and Fagella [9, Section 3.4] for further details. For semigroups, (see [15, section 5]) the authors show that if S acts on $\overline{\mathbb{C}}$, has disjoint limit sets, and is generated by finitely many non-elliptic maps, then S enjoys a property analogous to hyperbolicity for

rational maps. Specifically, there is an open set containing Λ^- and a metric defined on this set, upon which each element in S is expanding. The metric they define is indeed smooth and conformal. Theorem 4.23 shows that the limit sets of such semigroups vary continuously as the generating set is perturbed, at least within some small neighbourhood.

Theorem 4.23 begs the question: What is the domain of stability of Λ^+ and Λ^- ? That is for which $\mathcal{F}_0 \in \text{Aut}(\mathbb{B}^3)^k$ is the map $\mathcal{F} \mapsto \Lambda^+(\langle \mathcal{F} \rangle)$ continuous in some neighbourhood of \mathcal{F}_0 ? Indeed, one can ask similar questions with respect to other properties on semigroups. For example, is the property of being semidiscrete and inverse free stable under perturbation of the generators; in other words is the set

$$\left\{ \mathcal{F} \in \text{Aut}(\mathbb{B}^3)^k \mid S \text{ is semidiscrete and inverse free} \right\}$$

open in $\text{Aut}(\mathbb{B}^3)^k$?

We can however deduce that the set

$$\left\{ \mathcal{F} \in \text{Aut}(\mathbb{B}^3)^k \mid S \text{ is semidiscrete and inverse free and } \Lambda^+(S) \cap \Lambda^-(S) = \emptyset \right\}$$

is open, for each positive integer k . To see this suppose $\mathcal{F} \in \text{Aut}(\mathbb{B}^3)^k$ is such that $\langle \mathcal{F} \rangle$ has disjoint limit sets and is semidiscrete and inverse free. Then $\langle \mathcal{F} \rangle$ contains no elliptic elements or the identity, and so by Theorem 4.23 we can perturb the generators in some neighbourhood of \mathcal{F} and preserve the property that the semigroup they generate has disjoint limit sets. If the perturbation is sufficiently small, then the generators will not contain elliptic maps or the identity, and so by Theorem 4.16 they must generate a semidiscrete and inverse free semigroup.

We can still say something about how the limit sets vary with \mathcal{F} , even if the limit sets of S are not disjoint, as the next result shows. Throughout the proof of the next theorem it is convenient to adjust our notation slightly, and denote the attracting fixed point of a loxodromic Möbius transformation f by $\alpha(f)$ instead of α_f . For a set $X \subseteq \mathbb{S}^2$ and $\epsilon > 0$ we let $(X)_\epsilon$ denote the set of points in \mathbb{S}^2 whose chordal distance from any point in X is less than ϵ .

THEOREM 4.24. *Suppose $k \in \mathbb{N}$ and let $\mathcal{F} \in \text{Aut}(\mathbb{B}^3)^k$ be a bounded generating sequence generating a nonelementary semigroup. Then for all $\epsilon > 0$ there exists an open neighbourhood U of \mathcal{F} such that for all $\mathcal{G} \in U$,*

$$\Lambda^+(\langle \mathcal{F} \rangle) \subseteq (\Lambda^+(\langle \mathcal{G} \rangle))_\epsilon.$$

PROOF. For each $\mathcal{G} \in \text{Aut}(\mathbb{B}^3)^k$ let $\pi_{\mathcal{G}}$ be the projection map from $k^{<\omega}$ to the associated Möbius transformation in $\langle \mathcal{G} \rangle$. That is, if $\mathcal{G} = (g_1, \dots, g_k)$ then

$$\pi_{\mathcal{G}} : k^{<\omega} \rightarrow \text{Aut}(\mathbb{B}^3)$$

is given by

$$\pi((i_0, \dots, i_{N-1})) = g_{i_0} \cdots g_{i_{N-1}}$$

for each $(i_0, \dots, i_{N-1}) \in k^{<\omega}$. Since $\langle \mathcal{F} \rangle$ is nonelementary its set of attracting fixed points is dense in its forward limit set, and so we can choose a finite set of attracting fixed points such that the open $\epsilon/2$ -balls centred at these points cover Λ^+ . For each of these attracting fixed points α , choose a word $t \in k^{<\omega}$ such that $\pi_{\mathcal{F}}(t)$ has attracting fixed point α . We let T_ϵ denote the set of all such t , and so we have

$$\Lambda^+(\langle \mathcal{F} \rangle) \subseteq \bigcup_{t \in T_\epsilon} B(\alpha(\pi_{\mathcal{F}}(t)), \epsilon/2).$$

Since T_ϵ is finite, there is an open neighbourhood of \mathcal{F} , U say, such that for all $\mathcal{G} \in U$ and all $t \in T_\epsilon$ the Möbius transformation $\pi_{\mathcal{G}}(t)$ is loxodromic, and the point $\alpha(\pi_{\mathcal{G}}(t))$ is no further than $\epsilon/2$ from $\alpha(\pi_{\mathcal{F}}(t))$. This is possible because the collection of loxodromic transformations is open in $\text{Aut}(\mathbb{B}^3)$, and the attracting fixed point of a loxodromic Möbius transformation f varies continuously as f does. It follows that $B(\alpha(\pi_{\mathcal{F}}(t)), \epsilon/2) \subseteq B(\alpha(\pi_{\mathcal{G}}(t)), \epsilon)$ for all $t \in T_\epsilon$. Now since $B(\alpha(\pi_{\mathcal{G}}(t)), \epsilon) \subseteq (\Lambda^+(\langle \mathcal{G} \rangle))_\epsilon$ it follows that

$$\Lambda^+(\langle \mathcal{F} \rangle) \subseteq (\Lambda^+(\langle \mathcal{G} \rangle))_\epsilon$$

as required. □

The conclusion of the theorem can be regarded as ‘one half’ of continuity. For suppose the following statement were true: for all $\epsilon > 0$ there exists some neighbourhood U of \mathcal{F}

such that for all $\mathcal{G} \in U$,

$$(16) \quad \Lambda^+(\langle \mathcal{G} \rangle) \subseteq (\Lambda^+(\langle \mathcal{F} \rangle))_\epsilon.$$

When combined with the conclusion of the theorem, we could then infer that the map from $\text{Aut}(\mathbb{B}^3)^k$ to the space of compact subsets of \mathbb{S}^2 given by

$$\mathcal{F} \mapsto \Lambda^+(\langle \mathcal{F} \rangle)$$

is continuous. In fact condition (16) is false in general, as the example following Theorem 4.23 demonstrates.

3.3. Uniqueness of generating sets. Let $S = \langle \mathcal{F} \rangle$ where \mathcal{F} is a finite set. This subsection addresses the question of when \mathcal{F} is uniquely determined by S . We say \mathcal{F} is a *minimal generating set* if it has no redundant elements, that is, there are no elements we can remove from \mathcal{F} and still generate S from the remaining elements. We first give sufficient conditions that ensure S has a unique minimal generating set.

THEOREM 4.25. *Suppose S is generated by the finite minimal generating set \mathcal{F} , and has empty group part. Then \mathcal{F} is the unique minimal generating set if and only if the semigroup fSf^{-1} does not meet S^{-1} for all $f \in \mathcal{F}$.*

PROOF. First suppose S has two different minimal generating sets for S , say $\mathcal{F} = \{f_1, \dots, f_m\}$ and $\mathcal{G} = \{g_1, \dots, g_n\}$. Then some element of \mathcal{F} , which we take to be f_1 , does not lie in \mathcal{G} . Hence we can write f_1 as a word $g_{i_1} \dots g_{i_k}$ in \mathcal{G} with length $k \geq 2$. Now since each g_{i_j} lies in S , we can substitute g_{i_j} for some word in \mathcal{F} . We now have written f_1 as a word in \mathcal{F} , which we denote by w , and has length at least two. It is important to emphasise that no cancellations have been carried out; we regard words merely as sequences of letters, rather than sequences of Möbius transformations. If f_1 is the first or last letter of w , then f_1 can be cancelled from both sides leaving a nonempty word in \mathcal{F} equal to the identity, which contradicts that $I \notin S$. Yet if the letter f_1 does not feature as a letter in w , then $\{f_1, \dots, f_m\}$ is not a minimal generating set, contrary to assumption. Hence we must have $f_1 = w_1 f_1 w_2$ where w_1, w_2 are nonempty words in \mathcal{F} . Hence $w_1^{-1} = f_1 w_2 f_1^{-1}$ so that $f_1 S f_1^{-1}$ meets S^{-1} .

For the other direction, suppose fSf^{-1} meets S^{-1} for some $f \in \mathcal{F}$, say $f_1w_2f_1^{-1} = w_1^{-1}$ where w_1 and w_2 also lie in S . Hence $f_1 = w_1f_1w_2$. Then $\{w_1f_1, w_2, f_2, \dots, f_m\}$ is a generating set that does not contain f_1 . Hence any minimal generating set contained in $\{w_1f_1, w_2, f_2, \dots, f_m\}$ is a minimal generating set that does not contain f_1 . \square

The proof of the next corollary sheds light on the property $fSf^{-1} \cap S^{-1} = \emptyset$ for all $f \in \mathcal{F}$, and the corollary itself serves as further evidence that supposing a semigroup's limit sets are disjoint is a strong property on semigroups.

COROLLARY 4.26. *Suppose S is generated by a finite minimal generating set \mathcal{F} that contains no elliptic transformations. If the limit sets of S are disjoint then \mathcal{F} is a unique minimal generating set.*

PROOF. That S is semidiscrete and inverse free follows from Theorem 4.16, and so no element in S is elliptic. If \mathcal{F} were not a unique generating set, then fSf^{-1} would meet S^{-1} , say $fgf^{-1} = h^{-1}$. If g is loxodromic then so is h , and moreover f maps the axis of g to the axis of h , sending α_g to β_h and β_g to α_h . In particular the limit sets of S meet contrary to assumption. If g is parabolic then again, the limit sets of S meet, contrary to assumption. As g cannot be elliptic, \mathcal{F} must be a unique generating set after all. \square

We do not know if this implication runs in the other direction; however we recall the example given in Section 3 of Chapter 2 which is consistent with the converse.

In [15, Section 5] the authors show that whenever S is generated by a finite set \mathcal{F} acting on \mathbb{B}^3 such that each generator is not elliptic, then \mathcal{F} may be regarded as a contracting iterated function system acting on Λ^+ . They do this by considering the finitely many connected components of Ω^- that meet Λ^+ , and, as the limit sets are disjoint, these cover Λ^+ . They then construct a metric derived from the hyperbolic metrics on each of these components. In general S is not contracting on Λ^+ with respect to the chordal metric; however we can use Lemma 4.22 and ideas from the proof of Theorem 4.23 to find a finite set $\mathcal{F}_M \subseteq S$ such that $S_M = \langle \mathcal{F}_M \rangle$ has the same forward limit set as S , and \mathcal{F}_M is a contracting iterated function system with respect to the chordal metric, with limit set Λ^+ . To do this, we recall that each loxodromic $f \in \text{Aut}(\mathbb{B}^3)$ maps the interior of

$\mathbb{S}^2 \setminus I^+(f)$ onto $I^-(f)$. Crucially each f is contracting with respect to the chordal metric in the interior of $\mathbb{S}^2 \setminus I^+(f)$. By Lemma 4.22 each $f \in \mathcal{F}_M$ has small isometric discs (of radius less than $1/\sinh[\frac{1}{2}M]$). Since $\beta_f \in I^+(f)$, the disc $I^+(f)$ is close to Λ^- ; similarly $\alpha_f \in I^-(f)$. Hence for M large enough, no $I^-(f)$ meets $I^-(g)$ for any $f, g \in \mathcal{F}_M$. It follows that \mathcal{F}_M is contracting on

$$\mathbb{S}^2 \setminus \bigcup_{f \in \mathcal{F}_M} I^+(f)$$

and has forward limit set $\Lambda^+(S)$.

Collecting results from this chapter, we list some further properties of the class of semigroup considered above.

THEOREM 4.27. *Let S be a semigroup generated by a finite set of Möbius transformations that are not elliptic, and such that the limit sets of S are disjoint. Then*

- (i) *S is discrete,*
- (ii) *S has empty group part,*
- (iii) *both $\Lambda^+(S)$ and $\Lambda^-(S)$ vary continuously in a neighbourhood of \mathcal{F} ,*
- (iv) *S has a unique minimal generating set, and*
- (v) *every point in $\Lambda^+(S)$ is the limit point of some composition sequence, which converges conically.*

PROOF. By Theorem 4.16, S is semidiscrete and inverse free, hence by Theorem 4.21 S is discrete. Now (iii) and (iv) follow by Theorem 4.23 and Corollary 4.26 respectively. Finally (v) follows by Theorem 4.18. \square

CHAPTER 5

Self-maps of the disc

1. Introduction

This chapter is about composition sequences generated by finite subsets of the class $\text{Möb}_1(\mathbb{D})$ of those Möbius transformations that map the unit disc strictly within itself. If F_1, F_2, \dots is such a sequence, then the discs

$$\mathbb{D} \supseteq F_1(\mathbb{D}) \supseteq F_2(\mathbb{D}) \supseteq \dots$$

are nested, and the intersection $\bigcap F_n(\overline{\mathbb{D}})$ is either a single point or a closed disc (by *disc* we mean a disc of positive radius in $\overline{\mathbb{C}}$, using the chordal metric). In the first case we say that the composition sequence is of *limit-point type*, and in the second case we say that it is of *limit-disc type*. Composition sequences have been much studied within the literature on continued fractions (see, for example, [1, 4, 5, 27, 28]) – and the distinction between limit-point type and limit-disc type is of central importance. We will see that, for such composition sequences, limit-disc type really is quite special, and we find necessary and sufficient conditions for limit-disc type to occur. It is already known (see for example [27]) that every composition sequence of the type considered here converges ideally. Here we prove more, that every composition sequence has a strong convergence property. A composition sequence is of limit-point type exactly when the composition sequence converges uniformly on the unit disc. If the composition sequence is of limit-disc type, then the composition sequence merely converges locally uniformly (and not uniformly) on the unit disc.

To state our first result we introduce a new concept: a composition sequence F_n generated by a subset of $\text{Möb}_1(\mathbb{D})$ is said to be of *limit-tangent type* if all but a finite number of discs from the sequence $\mathbb{D} \supseteq F_1(\mathbb{D}) \supseteq F_2(\mathbb{D}) \supseteq \dots$ share a single common boundary point.

THEOREM 5.1. *Let \mathcal{F} be a finite subset of $\text{Möb}_1(\mathbb{D})$. Any composition sequence generated by \mathcal{F} of limit-disc type is of limit-tangent type.*

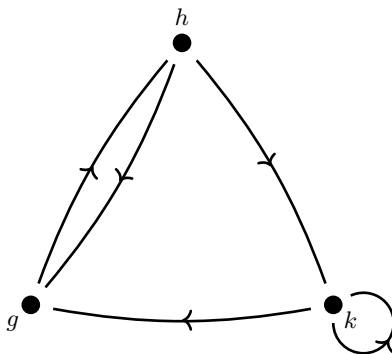
That a composition sequence F_n is of limit-tangent type implies that there is a point p and a sequence z_1, z_2, \dots in $\partial\mathbb{D}$ such that $p = F_n(z_n)$ for sufficiently large values of n . It follows that $f_n(z_n) = z_{n-1}$ for large n . We know that $f_n(\mathbb{D}) \subsetneq \mathbb{D}$, so we deduce that $f_n(\mathbb{D})$ is internally tangent to \mathbb{D} at the point z_{n-1} . Let us now choose any element f of $\text{Möb}_1(\mathbb{D})$: the disc $f(\mathbb{D})$ is either internally tangent to \mathbb{D} at a unique point, or else it does not touch the boundary of \mathbb{D} . In the former case, we define u_f and v_f to be the unique points in $\partial\mathbb{D}$ such that $f(u_f) = v_f$, and in the latter case, we define $u_f = 0$ and $v_f = \infty$ (for reasons of convenience to emerge shortly). The preceding theorem tells us that in order for the sequence F_n to be of limit-disc type, we need $u_{f_{n-1}} = v_{f_n}$ for sufficiently large values of n . How likely one is to find such a composition sequence is best illustrated by means of a directed graph $T(\mathcal{F})$, which we call the *tangency graph* of \mathcal{F} , and which is defined as follows. The vertices of $T(\mathcal{F})$ are the elements of \mathcal{F} . There is a directed edge from vertex f to vertex g if $u_f = v_g$. Clearly, any vertex f for which $f(\mathbb{D})$ and \mathbb{D} are not internally tangent is an isolated vertex.

Let us consider an example. One can easily check that a Möbius transformation $f(z) = a/(b+z)$, where $a \neq 0$, has the property that $f(\mathbb{D})$ is contained within and internally tangent to \mathbb{D} if and only if $|b| = 1 + |a|$. Let

$$g(z) = \frac{\frac{1}{2}}{\frac{3}{2} + z}, \quad h(z) = \frac{\frac{1}{2}}{-\frac{3}{2} + z}, \quad \text{and} \quad k(z) = \frac{-\frac{1}{2}}{-\frac{3}{2} + z}.$$

Each of these maps satisfies the condition $|b| = 1 + |a|$. Observe that $g(-1) = 1$, $h(1) = -1$ and $k(1) = 1$, which implies that $u_g = -1$, $v_g = 1$, $u_h = 1$, and so forth. The tangency graph of $\{g, h, k\}$ is shown in Figure 5.1.

From the tangency graph we can see that a composition sequence $F_n = f_1 \cdots f_n$ generated by $\{g, h, k\}$ is *not* of limit-tangent type if and only if (f_n, f_{n+1}) is equal to either (g, g) , (h, h) , (g, k) , or (k, h) for infinitely many positive integers n .

FIGURE 5.1. The tangency graph of $\{g, h, k\}$

We have shown that in order for a composition sequence F_n to be of limit-disc type, the sequence of vertices f_1, f_2, \dots in the tangency graph must eventually form an infinite path. The converse fails though: not all infinite paths in the tangency graph correspond to sequences of limit-disc type. To determine which paths arise from composition sequences of limit-disc type, we need to look at the derivatives of the maps f_n at tangency points. Let $\gamma_f = 1/|f'(u_f)|$, where $f \in \text{Möb}_1(\mathbb{D})$. The next theorem gives necessary and sufficient conditions for a composition sequence to be of limit-disc type.

THEOREM 5.2. *Let \mathcal{F} be a finite subcollection of $\text{Möb}_1(\mathbb{D})$. A composition sequence $F_n = f_1 \cdots f_n$ generated by \mathcal{F} is of limit-disc type if and only if:*

- (i) $u_{f_{n-1}} = v_{f_n}$ for all but finitely many positive integers n ; and
- (ii) $\sum_{n=1}^{\infty} \gamma_{f_1} \cdots \gamma_{f_n} < +\infty$.

For the example illustrated in Figure 5.1, one can check that $\gamma_g = \gamma_h = \gamma_k = \frac{1}{2}$. It follows that any composition sequence generated by $\{g, h, k\}$ satisfies condition (ii) of Theorem 5.2, so such a sequence is of limit-disc type if and only if it is of limit-tangent type.

Next we turn to the following question: given a finite subcollection \mathcal{F} of $\text{Möb}_1(\mathbb{D})$, how many of the composition sequences generated by \mathcal{F} are of limit-disc type? Using the tangency graph we can formalise this question, and answer it, at least in part, using known techniques. More precisely, we will define a metric on the set \mathcal{F}^ω of all sequences

taking values in \mathcal{F} . Each $x \in \mathcal{F}^\omega$ gives rise to a composition sequence, the sequence with n^{th} term $x(1) \cdots x(n)$.

Suppose (X, d) is any metric space and Y is a subset of X . We let $|Y|$ denote the diameter of Y . If $\{V_i\}$ is some finite or countable family of sets where $V_i \subseteq X$ and $\bigcup_i V_i$ contains Y , then we say the family $\{V_i\}$ is a *cover* of Y . If moreover $|V_i| \leq \delta$ for each i , then we say $\{V_i\}$ is a δ -*cover* of Y . Recall that (see for example [13, Chapter 2]) if $s \geq 0$ then we define the quantity

$$\mathcal{H}_\delta^s(Y) = \inf \left\{ \sum_i |U_i|^s \mid \{U_i\} \text{ is a } \delta\text{-cover of } Y \right\},$$

and define the *Hausdorff s -measure* of Y as

$$\inf_{\delta \rightarrow 0} \mathcal{H}_\delta^s(Y).$$

The *Hausdorff dimension* of Y is defined by

$$\dim(Y) = \inf \{s \geq 0 \mid \mathcal{H}^s(Y) = 0\} = \sup \{s \geq 0 \mid \mathcal{H}^s(Y) = +\infty\}.$$

We evaluate the Hausdorff dimension of $\Lambda_\circ(\mathcal{F})$, defined to be the set of elements in \mathcal{F}^ω whose associated composition sequences are of limit-disc type. The value of the Hausdorff dimension depends on the metric we choose; a standard choice for a metric on the set of sequences from a finite alphabet (which we adopt) is as follows. Given sequences x and y in \mathcal{F}^ω , we define

$$d(x, y) = \frac{1}{|\mathcal{F}|^m}, \quad \text{where } m = \min\{i \in \mathbb{N} \mid x(i) \neq y(i)\}.$$

Then (\mathcal{F}^ω, d) is a compact metric space.

A *cycle* in a directed graph is a sequence x_0, \dots, x_n of vertices in the graph such that $x_0 = x_n$ and there is a directed edge from x_{i-1} to x_i , for $i = 1, \dots, n$. Recall that the *spectral radius* of a square matrix is the largest modulus of the eigenvalues of the matrix. Let $\rho(\mathcal{F})$ denote the spectral radius of the adjacency matrix of $T(\mathcal{F})$. We can now state our first theorem on Hausdorff dimension.

THEOREM 5.3. *Let \mathcal{F} be a finite subcollection of $\text{Möb}_1(\mathbb{D})$, and suppose that $\gamma_f < 1$ for $f \in \mathcal{F}$. If \mathcal{F} contains a cycle, then*

$$\dim \Lambda_o(\mathcal{F}) = \frac{\log \rho(\mathcal{F})}{\log |\mathcal{F}|},$$

and otherwise $\dim \Lambda_o(\mathcal{F}) = 0$.

Returning again to the example illustrated in Figure 5.1, one can check that the spectral radius of the adjacency matrix of the tangency graph of $\mathcal{G} = \{g, h, k\}$ is $\frac{1}{2}(1 + \sqrt{5})$. Also, $\gamma_g, \gamma_h, \gamma_k < 1$. Therefore $\dim \Lambda_o(\mathcal{G}) = (\log(1 + \sqrt{5}) - \log 2) / \log 3$.

As we will see, Theorem 5.3 is a corollary of Theorem 5.2 and a well-known result on Hausdorff dimensions of paths in directed graphs. Also, because it assumes that condition (ii) from Theorem 5.2 is satisfied, it is really about the Hausdorff dimension of the set of composition sequences of limit-*tangent* type. In contrast, the next theorem determines $\dim \Lambda_o(\mathcal{F})$ under the assumption that condition (i) from Theorem 5.2 is satisfied.

THEOREM 5.4. *Let \mathcal{F} be a finite subcollection of $\text{Möb}_1(\mathbb{D})$, and suppose that $T(\mathcal{F})$ is the complete graph on $|\mathcal{F}|$ vertices. Then $\dim \Lambda_o(\mathcal{F}) = 0$ if $\gamma_f \geq 1$ for all $f \in \mathcal{F}$; $\dim \Lambda_o(\mathcal{F}) = 1$ if $\gamma_f < 1$ for all $f \in \mathcal{F}$; and otherwise*

$$\dim \Lambda_o(\mathcal{F}) = \frac{\log \left(\min \left\{ \sum_{f \in \mathcal{F}} \gamma_f^{-s} \mid s \geq 0 \right\} \right)}{\log |\mathcal{F}|}.$$

The problem of finding $\dim \Lambda_o(\mathcal{F})$ for an arbitrary finite subset \mathcal{F} of $\text{Möb}_1(\mathbb{D})$ remains open.

So far we have focused on determining those composition sequences of limit-disc type, and we have ignored questions of convergence of sequences in the unit disc. In fact, it is known already that any composition sequence generated by a finite subset of $\text{Möb}_1(\mathbb{D})$ converges locally uniformly in \mathbb{D} to a constant (this follows, for example, from [27, Theorems 3.8 and 3.10]). Here we will prove some stronger theorems on convergence of composition sequences generated by a finite set.

Recall that a sequence of Möbius transformations F_1, F_2, \dots is called an *escaping sequence* if $\rho(\zeta, F_n(\zeta)) \rightarrow \infty$ as $n \rightarrow \infty$ for any choice of the point ζ in \mathbb{H}^3 . We say that an escaping sequence F_1, F_2, \dots is of *convergence type* if $\sum \exp(-\rho(\mathbf{j}, F_n(\mathbf{j}))) < +\infty$. This terminology

is borrowed from the theory of Kleinian groups, and indeed, a sum of the type above will be familiar to Kleinian group theorists: it relates to the critical exponent of a Kleinian group, a connection that has been explored before in [35].

Our first theorem on convergence demonstrates that composition sequences generated by finite subsets of $\text{Möb}_1(\mathbb{D})$ are of convergence type. In fact, it is more general than this, as it does not assume that the generating set is finite. Let $\text{rad}(D)$ denote the Euclidean radius of a Euclidean disc D .

THEOREM 5.5. *Let f_1, f_2, \dots be a sequence of Möbius transformations such that $f_n(\mathbb{D}) \subseteq \mathbb{D}$ and $\text{rad}(f_n(\mathbb{D})) < \delta$, for $n = 1, 2, \dots$, where δ satisfies $0 < \delta < 1$. Then $F_n = f_1 \cdots f_n$ is of convergence type.*

The convergence type condition is a strong convergence property. By placing further, relatively mild conditions on sequences of convergence type (such as assuming that the sequence is generated by a finite set) one can prove that the sequence converges ideally. What is more, if F_n is of convergence type and converges ideally to a point p , then one can show that the set of points z in $\overline{\mathbb{C}}$ such that $F_n(z)$ does not converge to p as $n \rightarrow \infty$ has Hausdorff dimension at most 1 (see [35, Corollary 3.5]). For composition sequences generated by a finite set and of the type that interest us, we have the following result.

THEOREM 5.6. *Let $F_n = f_1 \cdots f_n$ be a composition sequence generated by a finite subcollection \mathcal{F} of $\text{Möb}_1(\mathbb{D})$. Then F_n converges ideally to a point p in $\overline{\mathbb{D}}$. Furthermore, F_n converges locally uniformly to p on $\mathbb{D} \setminus X$, where*

$$X = \{u_f \in \partial\mathbb{D} \mid f_n = f \text{ for infinitely many integers } n\}.$$

Theorem 5.6 is just a corollary of Theorem 5.5, and the part about converging ideally is known from other more general results (again, we refer the reader to [27, Theorems 3.8 and 3.10]). What we wish to emphasise is, first, the complete understanding of the dynamics that we have obtained, and, second, the interaction between the action of the transformations in hyperbolic space and on its boundary, which is illuminated in the proofs of the above pair of theorems. Although the results in this chapter are framed in terms of the unit disc, by conjugation they hold for any disc on the Riemann sphere. Indeed our results readily generalise to higher dimensions.

2. Eventually tangent sequences of discs

In this section we prove Theorem 5.1. We make the elementary observation that if distinct discs D and E satisfy $D \subseteq E$, then they can have at most one point of tangency.

LEMMA 5.7. *Let $D_1 \subseteq D_2 \subseteq \dots$ be a nested sequence of discs in the plane, no two of which are equal. The sequence is of limit-tangent type if and only if D_n is tangent to D_{n+2} for all but finitely many values of n .*

PROOF. Suppose that the sequence D_1, D_2, \dots is eventually tangent. Then there is a point p in the plane and a positive integer m such that $p \in \partial D_n$ when $n \geq m$. Therefore D_n and D_{n+2} share a common boundary point when $n \geq m$, and because the two discs are distinct and satisfy $D_{n+2} \subseteq D_n$, they must be tangent to one another.

Conversely, suppose that there is a positive integer m such that D_n is tangent to D_{n+2} when $n \geq m$. For any particular integer k , with $k \geq m$, let p be the point of tangency of D_k and D_{k+2} , and let c and r be the Euclidean centre and radius of D_{k+1} . As the discs D_1, D_2, \dots are nested, we see that first, $p \in \partial D_k$, so $|p - c| \geq r$, and second, $p \in \partial D_{k+2}$, so $|p - c| \leq r$. Therefore $|p - c| = r$, which implies that D_k , D_{k+1} and D_{k+2} have a (unique) common point of tangency. It follows that all the discs D_n have a common point of tangency when $n \geq m$. \square

We remark that the lemma remains true if we replace D_n and D_{n+2} with D_n and D_{n+q} , for any positive integer $q > 1$, but is false for $q = 1$.

PROOF OF THEOREM 5.1. We prove the contrapositive; that is, we prove that if $F_1(\mathbb{D}) \subseteq F_2(\mathbb{D}) \subseteq \dots$ is *not* eventually tangent, then F_1, F_2, \dots is of limit-point type. Suppose then that $F_1(\mathbb{D}) \subseteq F_2(\mathbb{D}) \subseteq \dots$ is not eventually tangent. By Lemma 5.7, there is an infinite sequence of positive integers $n_1 < n_2 < \dots$, where $n_{i+1} > n_i + 1$ for each i , such that $F_{n_i+2}(\mathbb{D})$ is not tangent to $F_{n_i}(\mathbb{D})$ for each integer n_i . Equivalently, $f_{n_i+1}f_{n_i+2}(\mathbb{D})$ is not tangent to \mathbb{D} for each integer n_i . Let K be the union of all the sets $fg(\overline{\mathbb{D}})$, where $f, g \in \mathcal{F}$, such that $fg(\mathbb{D})$ is not tangent to \mathbb{D} . Since \mathcal{F} is finite, K is a compact subset of \mathbb{D} .

Let us define $g_1 = f_1 \cdots f_{n_1}$ and $g_i = f_{n_i+1} \cdots f_{n_{i+1}}$ for $i = 2, 3, \dots$. Then, for $i > 1$,

$$g_i(\mathbb{D}) = f_{n_i+1} \cdots f_{n_{i+1}}(\mathbb{D}) \subseteq f_{n_i+1}f_{n_i+2}(\mathbb{D}) \subseteq K.$$

Let $G_n = g_1 \cdots g_n$. Since each g_i maps \mathbb{D} within a compact subset of \mathbb{D} , [4, Theorem 4.6] tells us that G_n converges uniformly on \mathbb{D} to a constant. That is, G_n is of limit-point type. Since the sequence $G_1(\mathbb{D}), G_2(\mathbb{D}), \dots$ is a subsequence of $F_1(\mathbb{D}), F_2(\mathbb{D}), \dots$, we deduce that F_n is also of limit-point type. \square

3. Necessary and sufficient conditions for limit-disc type

We define the *chordal derivative* of a Möbius transformation f by

$$f^\#(z) = \lim_{w \rightarrow z} \frac{\chi(f(z), f(w))}{\chi(z, w)}.$$

The chordal derivative is related to the Euclidean derivative (see [7, Equation 2.1]) by the equation

$$f^\#(z) = \frac{(1 + |z|^2)|f'(z)|}{1 + |f(z)|^2}$$

for $z \neq f^{-1}(\infty), \infty$.

In the ball model, the orientation-preserving isometries of \mathbb{B}^3 equipped with the Euclidean metric are exactly those Möbius transformations that fix $\mathbf{0}$. In the upper half-space model, the orientation preserving isometries of $(\overline{\mathbb{C}}, \chi)$ are exactly those Möbius transformations that fix \mathbf{j} .

Let us define

$$\gamma_f = \frac{1}{f^\#(u_f)},$$

where $f \in \text{Möb}_1(\mathbb{D})$. In fact, $f^\#(u_f)$ happens to coincide with $|f'(u_f)|$ (the Euclidean derivative), but we use the chordal derivative because it is defined everywhere on $\overline{\mathbb{C}}$, and could be used in a version of the next lemma in which \mathbb{D} is replaced by a half-plane (say). In this section we prove Theorem 5.2. Let \mathbb{K} denote the open right half-plane of \mathbb{C} and suppose that h is a Möbius transformation that satisfies $h(\mathbb{K}) \subseteq \mathbb{K}$ and $h(\infty) = \infty$. Then $h(z) = az + b$, where $a > 0$ and $\text{Re}[b] \geq 0$. Moreover, $h(\mathbb{K}) = \mathbb{K}$ if and only if $\text{Re}[b] = 0$.

LEMMA 5.8. *Suppose that h_n and H_n are sequences of Möbius transformations defined by $h_n(z) = a_n z + b_n$, where $a_n > 0$ and $\operatorname{Re}[b_n] > 0$, and $H_n = h_1 \cdots h_n$. The sequence of nested discs $H_1(\mathbb{K}), H_2(\mathbb{K}), \dots$ is of limit-disc type if and only if*

$$\sum_{n=1}^{\infty} a_1 \cdots a_{n-1} \operatorname{Re}[b_n] < +\infty.$$

We assume that the expression $a_1 \cdots a_{n-1}$ takes the value 1 when $n = 1$.

PROOF. Observe that $H_n(z) = a_1 \cdots a_n z + \sum_{k=1}^n a_1 \cdots a_{k-1} b_k$. Therefore

$$H_n(\mathbb{K}) = t_n + \mathbb{K}, \quad \text{where} \quad t_n = \sum_{k=1}^n a_1 \cdots a_{k-1} \operatorname{Re}[b_k];$$

the result follows immediately. \square

Whenever we have a composition sequence $F_n = f_1 \cdots f_n$, it will be convenient to define F_0 equal to I . We define Π to be the (hyperbolic) convex hull of \mathbb{D} in the upper half-space model, that is the smallest (hyperbolically) convex set in \mathbb{H}^3 containing each geodesic that lands at points in \mathbb{D} .

LEMMA 5.9. *Suppose that f_n is a sequence of Möbius transformations from $\operatorname{Möb}_1(\mathbb{D})$, and z_0, z_1, \dots is sequence of points from $\partial\mathbb{D}$ such that $f_n(z_n) = z_{n-1}$ for each positive integer n . Let $F_n = f_1 \cdots f_n$. Then the nested sequence of discs $F_1(\mathbb{D}), F_2(\mathbb{D}), \dots$ is of limit-disc type if and only if*

$$\sum_{n=1}^{\infty} \frac{\sinh \rho(\mathbf{j}, f_n(\Pi))}{F_{n-1}^{\#}(z_{n-1})} < +\infty.$$

PROOF. The discs \mathbb{D} and \mathbb{K} both have chordal radius 1, so there is a chordal isometry ϕ_n that maps \mathbb{D} to \mathbb{K} and maps z_n to ∞ , for $n = 0, 1, \dots$. Let $h_n = \phi_{n-1} f_n \phi_n^{-1}$ and $H_n = h_1 \cdots h_n$. Then $H_n = \phi_0 F_n \phi_n^{-1}$. Hence $H_n(\mathbb{K}) = \phi_0 F_n(\mathbb{D})$, so $F_1(\mathbb{D}), F_2(\mathbb{D}), \dots$ is of limit-disc type if and only if $H_1(\mathbb{K}), H_2(\mathbb{K}), \dots$ is of limit-disc type. Observe that $h_n(\infty) = \infty$ and $h_n \in \operatorname{Möb}_1(\mathbb{K})$. Therefore we can write $h_n(z) = a_n z + b_n$, where $a_n > 0$ and $\operatorname{Re}[b_n] > 0$. By applying the chain rule, and remembering that the maps ϕ_n are chordal isometries, we see that

$$F_n^{\#}(z_n) = H_n^{\#}(\infty) = \frac{1}{a_1 \cdots a_n}.$$

Next, since $h_n(\mathbb{K}) = b_n + \mathbb{K}$, a simple calculation in hyperbolic geometry shows that

$$\sinh \rho(\mathbf{j}, h_n(\Sigma)) = \operatorname{Re}[b_n],$$

where Σ is the hyperbolic plane with ideal boundary $\partial\mathbb{K}$. Observe that

$$\rho(\mathbf{j}, f_n(\Pi)) = \rho(\phi_{n-1}(\mathbf{j}), \phi_{n-1}f_n\phi_n^{-1}(\Sigma)) = \rho(\mathbf{j}, h_n(\Sigma)),$$

so $\sinh \rho(\mathbf{j}, f_n(\Pi)) = \operatorname{Re}[b_n]$. To conclude, we deduce from Lemma 5.8 that $F_1(\mathbb{D}), F_2(\mathbb{D}), \dots$ is of limit-disc type if and only if

$$\sum_{n=1}^{\infty} \frac{\sinh \rho(\mathbf{j}, f_n(\Pi))}{F_{n-1}^{\#}(z_{n-1})} < +\infty,$$

as required. \square

PROOF OF THEOREM 5.2. Statement (i) of Theorem 5.2 is equivalent to the assertion that the sequence of discs $F_1(\mathbb{D}) \subseteq F_2(\mathbb{D}) \subseteq \dots$ is eventually tangent. Therefore Theorem 5.1 tells us that if F_1, F_2, \dots is of limit-disc type, then statement (i) holds. Thus, we have only to show that, on the assumption that statement (i) holds, F_1, F_2, \dots is of limit-disc type if and only if statement (ii) holds.

Let $z_n = u_{f_n}$, for $n = 1, 2, \dots$, and $z_0 = v_{f_1}$. Then

$$f_n(z_n) = f_n(u_{f_n}) = v_{f_n} = u_{f_{n-1}} = z_{n-1},$$

for sufficiently large values of n . In particular, $z_k = f_{k+1} \cdots f_n(z_n)$, for $k = 1, \dots, n-1$.

Also, by the chain rule,

$$\gamma_1 \cdots \gamma_n = \frac{1}{f_1^{\#}(z_1) \cdots f_n^{\#}(z_n)} = \frac{1}{F_n^{\#}(z_n)}.$$

Since \mathcal{F} is finite, there is a positive constant k such that

$$\frac{1}{k} < \sinh \rho(\mathbf{j}, f(\Pi)) < k,$$

for any map f in \mathcal{F} . It follows that

$$\sum_{n=1}^{\infty} \gamma_{f_1} \cdots \gamma_{f_n} < +\infty \quad \text{if and only if} \quad \sum_{n=1}^{\infty} \frac{\sinh \rho(\mathbf{j}, f_n(\Pi))}{F_{n-1}^{\#}(z_{n-1})} < +\infty.$$

It follows then, from Lemma 5.9, that F_1, F_2, \dots is of limit-disc type if and only if statement (ii) holds, and this completes the proof. \square

4. Hausdorff dimension of the set of composition sequences of limit-disc type

In this section we prove Theorem 5.3. First we need some terminology, and a result [16, Corollary 2.9], which we state here, in our notation.

THEOREM 5.10. *Let G be a directed graph with vertices $\{0, \dots, b-1\}$ that contains a cycle. Then the Hausdorff dimension of the elements in $\{0, \dots, b-1\}^\omega$ that are paths in G is given by*

$$\frac{\log \rho(G)}{\log b},$$

where $\rho(G)$ is the spectral radius of G .

We say $x \in \mathcal{F}^\omega$ is a *path* in $T(\mathcal{F})$ if for each $n \in \mathbb{N}$ there is a directed edge in $T(\mathcal{F})$ from $x(n)$ to $x(n+1)$. We say $x \in \mathcal{F}^\omega$ is *eventually a path* in $T(\mathcal{F})$ if for some $m \in \mathbb{N}$ the sequence with n^{th} term $x(n+m)$ is a path in $T(\mathcal{F})$. We can now deduce Theorem 5.3.

PROOF OF THEOREM 5.3. By Theorem 5.2, a composition sequence is of limit-disc type if and only if the associated element in \mathcal{F}^ω is eventually a path in the tangency graph. That is, $\Lambda_o(\mathcal{F})$ is exactly the set of those elements in \mathcal{F}^ω that are eventually a path in $T(\mathcal{F})$. If $T(\mathcal{F})$ contains no cycles then clearly $\Lambda_o(\mathcal{F})$ is empty, and so has Hausdorff dimension 0. Otherwise we appeal to Theorem 5.10 and infer that the set of elements in \mathcal{F}^ω that are paths in $T(\mathcal{F})$ has Hausdorff dimension $\log \rho(\mathcal{F}) / \log b$. Hausdorff dimension enjoys a countable stability property, that is for any sequence of sets $X_n \subseteq \mathcal{F}^\omega$, we have $\dim \bigcup_{n=1}^\infty X_n = \sup_{n \in \mathbb{N}} \dim X_n$. As \mathcal{F} is finite, by this countable stability property it follows that the collection of elements in \mathcal{F}^ω that are eventually paths in $T(\mathcal{F})$ also has Hausdorff dimension $\log \rho(\mathcal{F}) / \log b$. \square

5. Complete tangency graph

In this section we turn to the proof of Theorem 5.4. Recall the metric d on \mathcal{F}^ω defined by $d(x, y) = 1/|\mathcal{F}|^m$ where $m = \min\{i \in \mathbb{N} \mid x(i) \neq y(i)\}$. So far we have worked with Hausdorff dimensions of subsets of the metric space (\mathcal{F}^ω, d) . By associating composition sequences generated by \mathcal{F} with sequences in \mathcal{F}^ω , we were able to assign a Hausdorff dimension to the collection of composition sequences that are of limit-disc type. In this section we represent composition sequences not by elements in \mathcal{F}^ω , but by points in $[0, 1]$,

via their base $|\mathcal{F}|$ -expansion. We think of composition sequences as points in the metric space $[0, 1]$ endowed with the Euclidean metric. In doing so we give rise to a different notion of Hausdorff dimension on sets of composition sequences. In fact, the two notions of Hausdorff dimension coincide. A form of this statement can be deduced from [14, Theorem 5.1], however for convenience we state and prove the result here, in our notation.

Let $b = |\mathcal{F}|$ and choose an enumeration $\{f_0, \dots, f_{b-1}\}$ of \mathcal{F} . We now regard sequences in \mathcal{F} as sequences in $\{0, \dots, b-1\}$. Define $I_1 \subseteq [0, 1]$ as the set of points whose base b -expansion is not eventually equal to 0, or eventually equal to $b-1$. Let $B_1 \subseteq \{0, \dots, b-1\}^\omega$ denote those sequences that are not eventually equal to 0, or eventually equal to $b-1$. Then when the projection $\pi : \{0, \dots, b-1\}^\omega \rightarrow [0, 1]$ given by $\pi(x) = \sum_{j=0}^{\infty} x(j)b^{-j}$ is restricted to B_1 , we obtain a bijection from B_1 onto I_1 . Given any $X \subseteq [0, 1]$ let $I^s(X)$ be the Hausdorff s -measure of X with respect to the Euclidean metric on $[0, 1]$. Similarly, for $Y \subseteq \{0, \dots, b-1\}^\omega$ let $B^s(Y)$ be the Hausdorff s -measure of Y induced by the metric space $(\{0, \dots, b-1\}^\omega, d)$. The quantities $I_\delta^s(X)$ and $B_\delta^s(Y)$ are defined in the obvious way. Since both $[0, 1] \setminus I_1$ and $\{0, \dots, b-1\}^\omega \setminus B_1$ are countable, $I^s(I_1) = B^s(B_1) = 1$ for all $s > 0$.

LEMMA 5.11. *For any $X \subseteq [0, 1]$ and $s > 0$ we have that $I^s(X) = 0$ if and only if $B^s(\pi^{-1}(X)) = 0$. Consequently the Hausdorff dimensions of X and $\pi^{-1}(X)$ are equal.*

PROOF. Suppose $X \subseteq I_1$ and choose any $s > 0$. Let us write $Y = \pi^{-1}(X)$. If $x, y \in Y$ are such that $d(x, y) = b^{-m}$, then $\pi(x)$ and $\pi(y)$ lie in the same b -adic interval of level $m-1$. Hence $|\pi(x) - \pi(y)| \leq b^{-m+1} = bd(x, y)$, and so π is a Lipschitz map. Hence by [14, Lemma 1.8] there exists $c > 0$ such that $I^s(\pi(Y)) \leq c^s B^s(Y)$ for all $Y \subseteq B_1$. In the other direction, suppose $Y \subseteq B_1$ and choose $\delta > 0$ and $\epsilon > 0$. Suppose $\{V_i\}$ is a family where each $V_i \subseteq I_1$, the family forms a δ -cover of $\pi(Y)$, and satisfies

$$\sum_i |V_i|^s \leq I_\delta^s(\pi(Y)) + \epsilon.$$

For each i let k_i be the unique positive integer such that

$$b^{-k_i-1} \leq |V_i| < b^{-k_i}.$$

Each V_i meets at most two b -adic intervals of level k_i . If V_i meets two such intervals we denote them by V_{i0} and V_{i1} . Otherwise V_i meets one such interval, which we denote by V_{i0} , and let $V_{i1} = \emptyset$. Note that

$$|\pi^{-1}(V_{ij})| \leq b^{-k_i-1}$$

so that

$$(17) \quad |\pi^{-1}(V_{ij})| \leq |V_i|.$$

Since $\{V_{ij}\}$ is a cover of $\pi(Y)$, the family $\{\pi^{-1}(V_{ij})\}$ is a cover of Y , and by (17) it is a δ -cover. Since $|V_{ij}| \leq b^{-k_i} \leq b|V_i|$ it follows that

$$\begin{aligned} B_\delta^s(Y) &\leq \sum_{i,j} |\pi^{-1}(V_{ij})|^s \leq \sum_{i,j} |V_{ij}|^s \leq 2b^s \sum_i |V_i|^s \\ &\leq 2b^s (I_\delta^s(\pi(Y)) + \epsilon). \end{aligned}$$

Letting $\epsilon \rightarrow 0$ we obtain $B_\delta^s(Y) \leq 2b^s I_\delta^s(\pi(Y))$. Hence

$$B_\delta^s(Y) \leq 2b^s I_\delta^s(\pi(Y)) \leq 2b^s I^s(\pi(Y)),$$

and letting $\delta \rightarrow 0$ yields $B^s(Y) \leq 2b^s I^s(\pi(Y))$.

Hence we have shown that

$$B^s(Y) \leq 2b^s I^s(\pi(Y)) \leq 2b^s c^s B^s(Y)$$

for all $Y \subseteq B_1$. Since $B \setminus B_1$ is countable, the above also holds for any $Y \subseteq B$. It follows that for any $X \subseteq [0, 1]$ we have $B^s(\pi^{-1}(X)) = 0$ if and only if $I^s(X) = 0$. \square

Above, we began by choosing an enumeration of \mathcal{F} in order to convert sequences in \mathcal{F} into sequences in $\{0, 1, \dots, b-1\}$. Because of the nature of the metric d on $\{0, \dots, b-1\}^\omega$, the Hausdorff dimension of subsets of $\{0, \dots, b-1\}^\omega$ is invariant under permutations of $\{0, 1, \dots, b-1\}$. It follows that the Hausdorff dimension we defined on composition sequences via the interval $[0, 1]$ is also independent of the choice of enumeration of \mathcal{F} .

We now work towards a proof of Theorem 5.4. Suppose $\{\gamma_0, \dots, \gamma_{b-1}\} \subseteq \mathbb{R}^+$. For each $x \in \{0, \dots, b-1\}^\omega$ and $i = 0, 1, \dots, b-1$ we write

$$N_i(x, n) = |\{j \in \{1, \dots, n\} \mid x(j) = i\}|.$$

In light of the above lemma and Theorem 5.2, our goal in this section is to prove the following.

THEOREM 5.12. *Suppose $\{\gamma_0, \dots, \gamma_{b-1}\} \subseteq \mathbb{R}^+$ is such that $\gamma_i > 1$ and $\gamma_j < 1$ for some i, j . Let X be the π -image of those $x \in \{0, \dots, b-1\}^\omega$ such that the series*

$$(18) \quad \sum_{n=1}^{\infty} \prod_{j=1}^n \gamma_{x(j)}$$

converges. Then

$$\dim X = \inf_{s>0} \log \left(\sum_{i=0}^{b-1} \gamma_i^{-s} \right) / \log b.$$

It is easy to find $\dim X$ if $\gamma_i \leq 1$ for all i , or if $\gamma_i \geq 1$ for all i . For example in the latter case, X is empty and $\dim X = 0$. Hence to prove Theorem 5.4 it suffices to prove Theorem 5.12. Given any $x \in \{0, \dots, b-1\}^\omega$ we define

$$\log \eta_n(x) = \sum_{i=0}^{b-1} \frac{N_i(x, n)}{n} \log \gamma_i,$$

and relate the convergence of the series

$$\sum_{n=1}^{\infty} \prod_{j=1}^n \gamma_{x(j)}$$

to the sequence $\eta_n(x)$.

LEMMA 5.13. *If $\limsup_{n \rightarrow \infty} \log \eta_n(x) < 0$ then the series*

$$\sum_{n=1}^{\infty} \prod_{j=1}^n \gamma_{x(j)}$$

converges, while, if $\limsup_{n \rightarrow \infty} \log \eta_n(x) > 0$, then the series diverges.

PROOF. This result is no more than an application of the n^{th} root test for converging series. It follows from the test that

$$(19) \quad \sum_{n=1}^{\infty} \prod_{j=1}^n \gamma_{x(j)}$$

converges if $\limsup_{n \rightarrow \infty} \left(\prod_{j=1}^n \gamma_{x(j)} \right)^{1/n} < 1$, but diverges if $\limsup_{n \rightarrow \infty} \left(\prod_{j=1}^n \gamma_{x(j)} \right)^{1/n} > 1$. Noting that the logarithm function is increasing, a straightforward computation shows that

$$\log \left(\frac{1}{n} \limsup_{n \rightarrow \infty} \prod_{j=1}^n \gamma_{x(j)} \right) = \limsup_{n \rightarrow \infty} \log \eta_n(x)$$

and so the result follows. \square

For $b = 1, 2, \dots$ and $x \in [0, 1]$ we write $I_n(x)$ as the n^{th} level b -adic half open interval containing x . We shall use the following result, which is also known as Billingsley's lemma, which can be found in [13, Proposition 2.3].

LEMMA 5.14. *Suppose $A \subseteq [0, 1]$ is a Borel set and let μ be a Borel measure on $[0, 1]$ such that $\mu([0, 1]) < \infty$. If for all $x \in A$ we have*

$$\liminf_{n \rightarrow \infty} \frac{\log \mu(I_n(x))}{\log |I_n(x)|} \leq s,$$

then $\dim A \leq s$. If $\mu(A) > 0$ and for all $x \in A$ we have

$$t \leq \liminf_{n \rightarrow \infty} \frac{\log \mu(I_n(x))}{\log |I_n(x)|},$$

then $t \leq \dim A$.

We say $p = (p_0, \dots, p_{b-1})$ is a *probability vector* if $0 \leq p_i \leq 1$ for each i , and $\sum_i p_i = 1$. For $i, j \in \{0, 1, \dots, b-1\}$ we define

$$\delta_{i,j} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

Using Lemma 5.14, the following result [8, Equation 14.4] can be shown.

LEMMA 5.15. *Let $p = (p_0, \dots, p_{b-1})$ be a probability vector. Given $x \in [0, 1]$ let $x(j)$ denote the j -th digit in the b -adic expansion of x . Let F_p be the set of $x \in [0, 1]$ such that for each $i = 0, \dots, b-1$*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \delta_{x(j), i} = p_i.$$

Then

$$\dim F_p = \frac{-\sum_i p_i \log p_i}{\log b}.$$

To prove Theorem 5.12 we first establish some properties of the function given by

$$g(s) = \log \sum_{i=0}^{b-1} \gamma_i^{-s}$$

on \mathbb{R} . For convenience we define $f(s) = -g'(s)$.

LEMMA 5.16. *Suppose that $\{\gamma_0, \dots, \gamma_{b-1}\} \subseteq \mathbb{R}^+$ satisfy the assumptions of Theorem 5.12. Then*

$$\lim_{s \rightarrow +\infty} g(s) = \lim_{s \rightarrow -\infty} g(s) = +\infty,$$

and g attains a unique global minimum in \mathbb{R} . The function f is strictly decreasing and has exactly one zero at the global minimum of g .

PROOF. The hypotheses on $\{\gamma_0, \dots, \gamma_{b-1}\}$ imply that $g(s) \rightarrow \infty$ as $|s| \rightarrow \infty$ does, and as g is continuous, $\{g(s) \mid s \in \mathbb{R}\}$ attains a minimum value $g(s_0)$ at some local minimum s_0 of g . Differentiating g gives

$$f(s) = \frac{\sum_i \gamma_i^{-s} \log \gamma_i}{\sum_i \gamma_i^{-s}}.$$

Note that as $s \rightarrow -\infty$ then for $\gamma > 1$, $\gamma^{-s} \log \gamma \rightarrow \infty$ while if $\gamma < 1$ then $\gamma^{-s} \log \gamma \rightarrow 0$. Similarly as $s \rightarrow +\infty$ then for $\gamma > 1$, $\gamma^{-s} \log \gamma \rightarrow 0$ while if $\gamma < 1$ then $\gamma^{-s} \log \gamma \rightarrow -\infty$. As the denominator of f is always positive, it follows that $f(s)$ takes negative values near $+\infty$ and positive values near $-\infty$, and in particular $f(s)$ has at least one zero. A computation shows that

$$\begin{aligned} f'(s) &= \frac{1}{(\sum_i \gamma_i^{-s})^2} \left[\left(\sum_i (\log \gamma_i) \gamma_i^{-s} \right)^2 - \left(\sum_i (\log \gamma_i)^2 \gamma_i^{-s} \right) \left(\sum_i \gamma_i^{-s} \right) \right] \\ &= f(s)^2 - \frac{\sum_i \gamma_i^{-s} (\log \gamma_i)^2}{\sum_i \gamma_i^{-s}}, \end{aligned}$$

so that $f'(s) < 0$ if $f(s) = 0$; that is, f is decreasing at its zeros. In fact f is a strictly decreasing function on \mathbb{R} : an application of the Cauchy-Schwarz inequality yields

$$\left(\sum_i (\log \gamma_i) \gamma_i^{-s} \right)^2 = \left(\sum_i (\gamma_i^{-s/2} \log \gamma_i) (\gamma_i^{-s/2}) \right)^2 \leq \left(\sum_i (\log \gamma_i)^2 \gamma_i^{-s} \right) \left(\sum_i \gamma_i^{-s} \right).$$

Moreover the inequality is strict, for otherwise $(\gamma_0^{-s/2}, \dots, \gamma_{b-1}^{-s/2})$ is a multiple of $(\gamma_0^{-s/2} \log \gamma_0, \dots, \gamma_{b-1}^{-s/2} \log \gamma_{b-1})$, which implies that $\log \gamma_i$ takes the same value over $i = 0, \dots, b-1$, contrary to our assumption on the γ_i . Since f is continuous it must have exactly one zero, so that $g(s)$ has exactly one local extremum in \mathbb{R} , which is hence also a global minimum. \square

Note that Lemma 5.16 says that the infimum of Theorem 5.12 is either attained as a minimum for some $s > 0$, or is equal to $g(0) = \log b$, in which case $\dim(X) = 1$.

Let us denote by s_0 the unique real number where g attains its minimum. We write

$$X_{<0} = \left\{ \pi(x) \mid x \in \{0, \dots, b-1\}^\omega \setminus E_2 \text{ and } \limsup_{n \rightarrow \infty} \log \eta_n(x) < 0 \right\},$$

and similarly define the set $X_{\leq 0}$, so that by Lemma 5.13 we have the inclusion

$$(20) \quad X_{<0} \subseteq X \subseteq X_{\leq 0}.$$

We can now prove Theorem 5.12.

PROOF OF THEOREM 5.12. For $s \in \mathbb{R}$ we define μ_s to be the probability measure generated by the probability vector

$$\frac{1}{\sum_i \gamma_i^{-s}} (\gamma_0^{-s}, \dots, \gamma_{b-1}^{-s}).$$

Now, for $x \in \{0, \dots, b-1\}^\omega$, let $I_n(x)$ denote the n -level b -adic interval containing $\pi(x)$.

It follows that

$$\begin{aligned} \frac{\log \mu_s(I_n(x))}{\log |I_n(x)|} &= \frac{\log \left((\sum_i \gamma_i^{-s})^{-n} \prod_i \gamma_i^{-s N_i(x, n)} \right)}{\log b^{-n}} \\ &= \frac{-n \log(\sum_i \gamma_i^{-s}) - s \sum_i N_i(x, n) \log \gamma_i}{-n \log b}, \\ (21) \quad \text{hence} \quad \liminf_{n \rightarrow \infty} \frac{\log \mu_s(I_n(x))}{\log |I_n(x)|} &= \frac{\log(\sum_i \gamma_i^{-s})}{\log b} + \liminf_{n \rightarrow \infty} \frac{s}{\log b} \log \eta_n(x). \end{aligned}$$

We now show that $\dim(X) \geq \inf_{s>0} \log(\sum_i \gamma_i^{-s}) / \log b$. Let F_s denote the set of $x \in [0, 1]$ such that for each $j = 0, \dots, b-1$

$$\lim_{n \rightarrow \infty} \frac{N_j(x, n)}{n} = \frac{\gamma_j^{-s}}{\sum_i \gamma_i^{-s}}.$$

If $x \in F_s$ then

$$\lim_{n \rightarrow \infty} \log \eta_n(x) = \frac{\sum_i \gamma_i^{-s} \log \gamma_i}{\sum_i \gamma_i^{-s}} = f(s).$$

By Lemma 5.16 we have $f(s_0 + \epsilon) < 0$ and so $F_{s_0+\epsilon} \subseteq X$ by Lemma 5.13. Now applying Lemma 5.15 gives

$$\dim(F_s) = \frac{g(s)}{\log b} + \frac{sf(s)}{\log b};$$

hence for all $\epsilon > 0$ we have

$$\dim(X) \geq \dim(F_{s_0+\epsilon}) = \frac{g(s_0 + \epsilon) + (s_0 + \epsilon)f(s_0 + \epsilon)}{\log b}.$$

If $s_0 < 0$ then set $\epsilon = -s_0$ to obtain $\dim(X) \geq g(0)/\log(b) = 1$, so that $\dim(X) = 1$. Otherwise $s_0 \geq 0$, and since $f(s_0 + \epsilon) \rightarrow 0$ as $\epsilon \rightarrow 0$, we obtain the bound

$$\dim(X) \geq \frac{g(s_0)}{\log b}.$$

Finally we give the upper bound on $\dim(X)$. We know that $\dim(X) = 1$ if $s_0 < 0$, so we need only consider $s_0 \geq 0$. Now if $s \geq 0$ and $x \in X_{\leq 0}$ then equation (21) gives

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{\log \mu_s(I_n(x))}{\log |I_n(x)|} &= \frac{g(s)}{\log b} + \frac{s}{\log b} \liminf_{n \rightarrow \infty} \log \eta_n(x) \\ &\leq \frac{g(s)}{\log b} + \frac{s}{\log b} \limsup_{n \rightarrow \infty} \log \eta_n(x) \\ &\leq \frac{g(s)}{\log b}. \end{aligned}$$

Now by Lemma 5.14 and by the set inclusion (20) we have

$$\dim(X) \leq \dim(X_{\leq 0}) \leq \frac{g(s)}{\log b}.$$

Since g attains its minimum at $s_0 \geq 0$ we obtain the bound

$$\dim(X) \leq \inf_{s>0} \frac{g(s_0)}{\log b}.$$

Amalgamating our findings, if $s_0 \geq 0$ then we have $\dim(X) = g(s_0)/\log b$, while if $s_0 \leq 0$, then $\dim(X) = g(0)/\log b = 1$.

□

6. Convergence of composition sequences

So far in this chapter we have concerned ourselves with the dichotomy between limit-point and limit-disc type. We now turn to the issue of convergence, and in particular we prove Theorem 5.5. Recall that the height of a point $z + t\mathbf{j} \in \mathbb{H}^3$ is denoted by $\text{ht}[z + t\mathbf{j}] = t$.

LEMMA 5.17. *Let U and V be distinct Euclidean discs with centres u and v , and radii r and s , in that order. Suppose that $V \subseteq U$ and*

$$(22) \quad \frac{r}{s} < \min \left\{ 2, 1 + \frac{1}{8} \sinh \rho(z, \Pi(V)) \right\},$$

for some point z in $\Pi(U)$. Let w be the point in $\Pi(V)$ with $\text{ht}[w] = \text{ht}[z]$ that is closest in Euclidean distance to z . Then

$$|z - w| < 2(r - s).$$

PROOF. Using a standard formula for the hyperbolic metric (see, for example, [3, Section 7.20]) we have

$$\sinh \rho(z, \Pi(V)) = \frac{|z - v|^2 - s^2}{2hs},$$

where $h = \text{ht}[z]$. As $V \subseteq U$, we see that $|u - v| + s \leq r$. Hence

$$|z - v|^2 - s^2 \leq (|z - u| + |u - v|)^2 - s^2 \leq (2r - s)^2 - s^2 = 4r(r - s).$$

Therefore, using (22), we obtain

$$\sinh \rho(z, \Pi(V)) \leq \frac{2r(r - s)}{hs} = \frac{2r}{h} \left(\frac{r}{s} - 1 \right) < \frac{r}{4h} \sinh \rho(z, \Pi(V)).$$

It follows that $h < r/4 < s/2$.

Now, let Σ denote the Euclidean plane in \mathbb{H}^3 with height h . Let U_0 and V_0 denote the Euclidean discs in Σ with centres $u + h\mathbf{j}$ and $v + h\mathbf{j}$, and radii $\sqrt{r^2 - h^2}$ and $\sqrt{s^2 - h^2}$, in that order. Observe that $V_0 \subseteq U_0$ and $z \in U_0$ and $w \in V_0$. Using elementary Euclidean geometry, we can see that

$$|z - w| < 2\sqrt{r^2 - h^2} - 2\sqrt{s^2 - h^2}.$$

But $h < s/2$ and $h < r/2$, so

$$2\sqrt{r^2 - h^2} - 2\sqrt{s^2 - h^2} = \frac{2(r^2 - s^2)}{\sqrt{r^2 - h^2} + \sqrt{s^2 - h^2}} < \frac{2}{\sqrt{3}}(r - s).$$

Therefore $|z - w| < 2(r - s)$, as required. \square

PROOF OF THEOREM 5.5. Let $\eta = \min \{2, 1 + \frac{1}{8} \sinh \delta\}$ and let r_n be the Euclidean radius of $F_n(\mathbb{D})$ (with $r_0 = 1$). Now define

$$A = \left\{ n \in \mathbb{N} \mid \frac{r_{n-1}}{r_n} < \eta \right\} \quad \text{and} \quad B = \left\{ n \in \mathbb{N} \mid \frac{r_{n-1}}{r_n} \geq \eta \right\}.$$

Suppose that $n \in A$. Observe that $\rho(F_{n-1}(\mathbf{j}), F_n(\Pi)) = \rho(\mathbf{j}, f_n(\Pi)) > \delta$, so

$$\frac{r_{n-1}}{r_n} < \eta < \min \left\{ 2, 1 + \frac{1}{8} \sinh \rho(F_{n-1}(\mathbf{j}), F_n(\Pi)) \right\}.$$

Let w_n be the point in $F_n(\Pi)$ with $\text{ht}[w_n] = \text{ht}[F_{n-1}(\mathbf{j})]$ that is closest in Euclidean distance to $F_{n-1}(\mathbf{j})$. Lemma 5.17 tells us that

$$|F_{n-1}(\mathbf{j}) - w_n| < 2(r_{n-1} - r_n).$$

Now, using a basic estimate of hyperbolic distance, we find that

$$\delta \text{ht}[F_{n-1}(\mathbf{j})] < \rho(F_{n-1}(\mathbf{j}), F_n(\Pi)) \text{ht}[F_{n-1}(\mathbf{j})] \leq \rho(F_{n-1}(\mathbf{j}), w_n) \text{ht}[F_{n-1}(\mathbf{j})] \leq |F_{n-1}(\mathbf{j}) - w_n|.$$

Hence $\delta \text{ht}[F_{n-1}(\mathbf{j})] < 2(r_{n-1} - r_n)$, so

$$\sum_{n \in A} \text{ht}[F_{n-1}(\mathbf{j})] < +\infty.$$

Next, let $n_1 < n_2 < \dots$ be the elements of B . As $r_{n_k}/r_{n_{k-1}} \leq 1/\eta$, we see that $r_{n_k} \leq 1/\eta^k$.

Therefore

$$\text{ht}[F_{n_k-1}(\mathbf{j})] \leq r_{n_k-1} \leq r_{n_{k-1}} \leq \frac{1}{\eta^k},$$

so

$$\sum_{n \in B} \text{ht}[F_{n-1}(\mathbf{j})] < +\infty.$$

Observe that $-\log \text{ht}[F_{n-1}(\mathbf{j})] \leq \rho(\mathbf{j}, F_{n-1}(\mathbf{j}))$, from which it follows that $\exp[-\rho(\mathbf{j}, F_{n-1}(\mathbf{j}))] \leq \text{ht}[F_{n-1}(\mathbf{j})]$. We have just seen that

$$\sum_{n=1}^{\infty} \text{ht}[F_{n-1}(\mathbf{j})] < +\infty,$$

and hence we conclude that F_1, F_2, \dots is of convergence type. \square

PROOF OF THEOREM 5.6. Let us begin by proving that F_n converges ideally. In this part of the proof it is convenient to switch to the unit ball model of hyperbolic space. In this model the distinguished point $\mathbf{0}$ replaces the point \mathbf{j} in \mathbb{H}^3 . Two standard formulas for the hyperbolic metric in \mathbb{B}^3 are

$$\rho(\mathbf{0}, z) = \log \frac{1 + |z|}{1 - |z|} \quad \text{and} \quad \sinh \frac{1}{2} \rho(z, w) = \frac{|z - w|}{\sqrt{(1 - |z|^2)(1 - |w|^2)}}.$$

From the first standard formula we see that $1 - |z| = (1 + |z|)e^{-\rho(\mathbf{0}, z)}$. Substituting this equation into the second standard formula, and noting that $1 + |z| \leq 2$, we obtain

$$|z - w| = \sqrt{(1 - |z|^2)(1 - |w|^2)} \sinh \frac{1}{2} \rho(z, w) \leq 2(e^{-\rho(\mathbf{0}, z)} + e^{-\rho(\mathbf{0}, w)}) \sinh \frac{1}{2} \rho(z, w).$$

We apply this formula with $z = F_{n-1}(\mathbf{0})$ and $w = F_n(\mathbf{0})$. Let $k > 0$ be such that $\rho(\mathbf{0}, f_n(\mathbf{0})) < k$ for every positive integer n (as there are only finitely many different maps f_n). Then $\rho(F_{n-1}(\mathbf{0}), F_n(\mathbf{0})) < k$ for each integer n . Therefore

$$|F_{n-1}(\mathbf{0}) - F_n(\mathbf{0})| < 2 \sinh \frac{1}{2} k (\exp(-\rho(\mathbf{0}, F_{n-1}(\mathbf{0}))) + \exp(-\rho(\mathbf{0}, F_n(\mathbf{0}))).$$

Theorem 5.5 tells us that F_n is of convergence type, so we can see that $\sum |F_{n-1}(\mathbf{0}) - F_n(\mathbf{0})| < +\infty$. It follows that $F_1(\mathbf{0}), F_2(\mathbf{0}), \dots$ converges in the Euclidean metric on $\overline{\mathbb{B}^3}$. This sequence cannot converge to a point in \mathbb{B}^3 , because F_n is an escaping sequence. Therefore F_n converges ideally.

Let us now go back to thinking about F_n acting on $\overline{\mathbb{C}}$ and \mathbb{H}^3 . We have seen that F_n converges ideally to a point p . This point p must belong to $\overline{\mathbb{D}}$ because the orbit $F_n(\mathbf{j})$ is constrained within the Euclidean interior of Π .

It remains to show that F_n converges locally uniformly to p on $\overline{\mathbb{D}} \setminus X$, where X is the finite set $X = \{u_f \in \partial\mathbb{D} \mid f_n = f \text{ for infinitely many integers } n\}$. To prove this, we first note that it follows from Theorem 1.1 that F_n converges locally uniformly to p on the complement of its backward limit set. In our circumstances, the backward limit set is contained in the complement of \mathbb{D} . Therefore F_n converges locally uniformly to p on \mathbb{D} .

Let us choose $q \in \overline{\mathbb{D}} \setminus X$. Then there is a compact subset K of \mathbb{D} such that $f(q) \in K$ for all $f \in \mathcal{F}$. By increasing the size of K slightly (still a compact subset of \mathbb{D}) we can assume that there is an open Euclidean disc Q centred on q that does not intersect X such

that $f(Q) \subseteq K$ for all $f \in \mathcal{F}$. It follows that F_n converges locally uniformly to p on the complement of X in $\overline{\mathbb{D}}$, as required. \square

CHAPTER 6

Coding limit sets

This contents of this chapter are work in progress, and the author expects that some results are likely to hold in a more general setting than the one described here.

Consider the semigroup generated by the two parabolic transformations $f_0(z) = z + 1$ and $f_1(z) = \frac{z}{z+1}$ in $\text{Aut}(\mathbb{H}^2)$. Then the forward limit set is equal to the interval $[0, \infty] \subseteq \overline{\mathbb{R}}$. Moreover for every irrational number x in $[0, \infty]$ there is a unique composition sequence generated by f_0 and f_1 that converges ideally to x . Indeed, any composition sequence that converges ideally to a point x is essentially the usual continued fraction expansion of x of the form

$$x = a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \cdots}}.$$

In this chapter we consider semigroups such that, for some points in the semigroup's forward limit set (and as we shall see, most points), there exists more than one composition sequence that converges ideally to that point. Both for ease of analysis and exposition we work with a narrow class of semigroup. We choose two loxodromic elements f_0 and f_1 in $\text{Aut}(\mathbb{D})$, such that the set of attracting fixed points $\{\alpha_{f_0}, \alpha_{f_1}\}$ does not meet the set of repelling fixed points $\{\beta_{f_0}, \beta_{f_1}\}$, and both repelling fixed points lie in the same component of $\mathbb{S}^1 \setminus \{\alpha_{f_0}, \alpha_{f_1}\}$. Let σ_v denote a reflection in the imaginary axis and let σ_h denote a reflection in the real axis. The axes of f_0 and f_1 either meet in \mathbb{D} , or not. In both cases we can conjugate the semigroup such that the axes of the generators enjoy the following symmetries: both axes $\text{ax}(f_0)$ and $\text{ax}(f_1)$ are fixed setwise by σ_h , and σ_v transposes $\text{ax}(f_0)$ and $\text{ax}(f_1)$. Examples of the two cases are shown in Figure 6.1. In both

cases the connected component of $\mathbb{S}^1 \setminus \{\alpha_{f_0}, \alpha_{f_1}\}$ that does not contain either β_{f_0} or β_{f_1} , shown in red, is mapped strictly within itself by both generators. Hence S is semidiscrete and inverse free by Theorem 2.5, and the forward limit set is contained in the red interval, by Theorem 1.8 applied to S^{-1} . Similarly it follows that, in both cases, the backward limit set of S is contained in the blue interval.

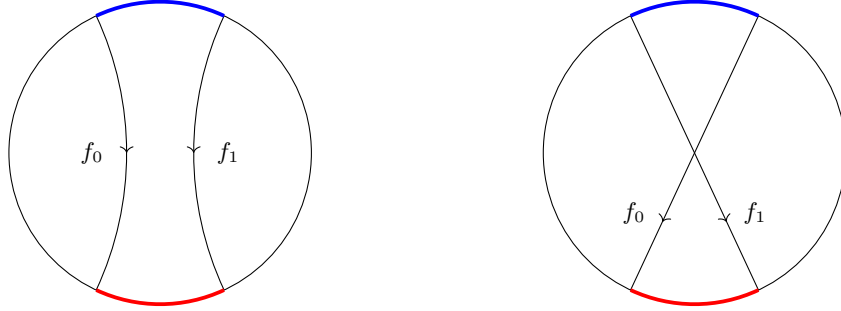


FIGURE 6.1. Possible configurations for the axes of f_0 and f_1

Let S be the semigroup generated by f_0 and f_1 . Let $\{0, 1\}^\omega$ denote the set of all sequences taking values in $\{0, 1\}$, and let $\{0, 1\}^n$ denote the set of all functions from $\{0, 1, \dots, n-1\}$ to $\{0, 1\}$. We write

$$\{0, 1\}^{<\omega} = \bigcup_{n=0}^{\infty} \{0, 1\}^n$$

as the set of all finite sequences taking values in $\{0, 1\}$. In a similar manner to the proof of Lemma 4.22, we endow $\{0, 1\}^{<\omega}$ with the usual partial order relation; that is for $i, j \in \{0, 1\}^\omega$ then $i \leq j$ if and only if the domain of i is contained in the domain of j and i agrees with j on the domain of i . Hence we can regard $\{0, 1\}^{<\omega}$ as a tree.

For any $i \in \{0, 1\}^{<\omega}$ with domain $\{0, 1, \dots, n-1\}$ we write $|i| = n$. For any $m \leq |i|$ we let $i \upharpoonright m = (i_0, \dots, i_{m-1})$ be the restriction of i to $\{0, 1, \dots, m-1\}$. For any $i \in \{0, 1\}^n$ we abbreviate the composition sequence $f_{i_0} \cdots f_{i_{n-1}}$ to F_i , and we define F_\emptyset to be the identity. For i and j in $\{0, 1\}^{<\omega}$ we let i, j denote the concatenation of i and j . We call the elements $(i_0, \dots, i_{|i|-1}, 0)$ and $(i_0, \dots, i_{|i|-1}, 1)$ in $\{0, 1\}^{<\omega}$ the *children* of $i = (i_0, \dots, i_{|i|-1})$.

Let X denote the closed interval in \mathbb{S}^1 bounded by the attracting fixed points of the generators that does not contain the generators' repelling fixed points. Since S is semidiscrete

and inverse free it follows from Theorem 4.18 that for any $i \in \{0, 1\}^\omega$ the composition sequence with n^{th} term $F_{i|n} = f_{i_0} \cdots f_{i_{n-1}}$ converges ideally to a point, which we shall denote by $x(i)$. It follows from Theorem 1.1 that the sequence $F_{i|n}(X)$ converges to the singleton $\{x(i)\}$. Indeed, $F_{i|n}(X)$ is a decreasing nest of sets since $F_{i|n}$ is a composition sequence. If for $i \in \{0, 1\}^\omega$ the sequence $F_{i|n}$ converges ideally to x , we say i is an *address* of x . Theorem 4.18 also tells us that each point in the forward limit set is the limit point of some composition sequence generated by \mathcal{F} . In other words, each point in Λ^+ has an address. Let Y denote the set $f_0(X) \cap f_1(X)$, and let us write $a = \alpha_{f_0}$ and $b = \alpha_{f_1}$. If Y is empty then each point in Λ^+ has a unique address, as is the case in the familiar Cantor semigroup (although, strictly speaking, the Cantor semigroup is not an example of the type of semigroup we consider here as its generators' repelling fixed points coincide). Since X is forward invariant under S , the forward limit set is certainly contained within X , and indeed $\Lambda^+ = X$ if and only if Y is not empty. As we shall refer to them frequently, for any $j \in \{0, 1\}^{<\omega}$ we abbreviate $F_j(X)$ as X_j and $F_j(Y)$ as Y_j .

In this chapter we focus on the case where Y has a nonempty interior, and so each point $x \in X$ has some address, which is not necessarily unique. We shall classify points in \mathbb{S}^1 by how many addresses they have. Clearly any point not in X has no address, while the points a and b always have exactly one address.

For a cardinal κ we define E_κ as the set of $x \in \mathbb{S}^1$ such that the set of addresses for x has cardinality κ . We write $E_{<\kappa}$ as the set of $z \in \mathbb{S}^1$ such that $\{i \in \{0, 1\}^\omega \mid x(i) = z\}$ has cardinality less than κ . The sets $E_{>\kappa}$ and $E_{\leq \kappa}$ are similarly defined. Our first goal is to show that the set of points in X with infinitely many addresses, that is $E_{\geq \aleph_0}$, is a 'topologically large set'. Towards this we first give another description of $E_{\geq \aleph_0}$. Recall that we say a branch of a tree is cofinal if it intersects every level of the tree.

LEMMA 6.1. *Let f_0 and f_1 be two loxodromic elements in $\text{Aut}(\mathbb{D})$ that are not antiparallel, and let X and Y be the sets described above, which depend on f_0 and f_1 . If Y is non-empty, then $\Lambda^+(\langle f_0, f_1 \rangle) = X$. If Y has a non-empty interior, then the set of points in X*

with infinitely many addresses is given by

$$\begin{aligned} E_{\geq \aleph_0} &= \{x \in X \mid \{i \in \{0,1\}^{<\omega} \mid F_i^{-1}(x) \in Y\} \text{ is infinite} \} \\ &= \bigcap_{n=0}^{\infty} \bigcup_{\substack{k=n \\ i \in \{0,1\}^k}}^{\infty} F_i(Y). \end{aligned}$$

PROOF. If Y is non-empty then $X = f_0(X) \cup f_1(X)$, so that for each $x \in X$ there exists $j \in \{0,1\}$ such that $x \in f_j(X)$. Hence we can recursively choose $i \in \{0,1\}^\omega$ such that $f_{i_{n-1}}^{-1} \cdots f_{i_1}^{-1} f_{i_0}^{-1}(x) \in X$, that is $x \in f_{i_0} f_{i_1} \cdots f_{i_{n-1}}(X)$ for each $n \in \mathbb{N}$. The composition sequence $f_{i_0} f_{i_1} \cdots f_{i_{n-1}} = F_{i|_n}$ converges ideally, and since X does not meet Λ^- this convergence is uniform on X . Since $x \in F_{i|_n}(X)$ for all n , it follows that $F_{i|_n}$ converges ideally to x , and in particular, x lies in the forward limit set of $\langle f_0, f_1 \rangle$. Hence $\Lambda^+(\langle f_0, f_1 \rangle) = X$ as claimed.

Now suppose that Y has non-empty interior. Given each $x \in X$ let $T(x)$ be the set of $i \in \{0,1\}^{<\omega}$ such that $F_i^{-1}(x) \in X$. $T(x)$ is a tree, since for $i \in T(x)$, $X_i \subseteq X_{i|_n}$ for each $n = 0, 1, \dots, |i|$. Furthermore, as $Y = f_0(X) \cap f_1(X)$, if i belongs to $T(x)$, then both its children also belong to $T(x)$ if and only if $F_i^{-1}(x) \in Y$. The tree $T(x)$ has infinitely many vertices since x has at least one address. If only finitely many vertices of $T(x)$ have two children, then as a subtree of $\{0,1\}^{<\omega}$, $T(x)$ has only finitely many cofinal branches. Conversely if infinitely many vertices in $T(x)$ have two children, then $T(x)$ has infinitely many cofinal branches. It follows that $E_{\geq \aleph_0} = \{x \in X \mid \{i \in \{0,1\}^{<\omega} \mid F_i^{-1}(x) \in Y\} \text{ is infinite} \}$. Finally we observe that

$$\bigcap_{n=0}^{\infty} \bigcup_{\substack{k=n \\ i \in \{0,1\}^k}}^{\infty} F_i(Y)$$

is simply a set-theoretic description of $\{x \in X \mid \{i \in \{0,1\}^{<\omega} \mid F_i^{-1}(x) \in Y\} \text{ is infinite} \}$. \square

In fact it can be shown that for any subtree of $\{0,1\}^\omega$, there are either finitely many, countably infinitely many or continuum many cofinal branches. For $x \in X$, the tree $T(x)$ from the proof of Lemma 6.1 is a subtree of $\{0,1\}^\omega$, and so x either has finitely, countably infinitely many or continuum many addresses. Hence E_κ is nonempty only for κ finite,

$\kappa = \aleph_0$ and $\kappa = 2^{\aleph_0}$.

Recall from Section 2.1 in Chapter 4 that a subset of a topological space is meagre if it is the countable union of nowhere dense sets. A residual set is the complement of a meagre set. Meagre sets are ‘topologically small’ sets – they form a σ -ideal, and so residual sets can be regarded as ‘topologically large’ sets.

COROLLARY 6.2. *Let f_0 and f_1 be two loxodromic elements in $\text{Aut}(\mathbb{D})$ that are not antiparallel, and let X be the set described above, that is the closed interval bounded by α_{f_0} and α_{f_1} that does not contain β_{f_0} or β_{f_1} . Then $E_{\geq \aleph_0}$ is a residual subset of X .*

PROOF. Since Y is a proper interval, $E_{\geq \aleph_0}$ contains the G_δ set

$$U = \bigcap_{n=0}^{\infty} \bigcup_{\substack{k=n \\ i \in \{0,1\}^k}}^{\infty} F_i(Y^\circ),$$

where Y° is the interior of Y . Since

$$\bigcup_{\substack{k=n \\ i \in \{0,1\}^k}}^{\infty} F_i(Y^\circ)$$

is open and dense in X , it follows from the Baire category theorem that U is dense in X , and so $E_{\geq \aleph_0}$ contains a dense G_δ subset of X . It follows that $E_{\geq \aleph_0}$ is a residual set. \square

Recall the chordal metric defined in Equation 1 in Chapter 5. We next record two distortion results, which can be found in [34, Equation 1.3.2] and [34, Theorem 1.3.4], that we shall need again.

LEMMA 6.3. *In the disc model, for any Möbius transformation g fixing the unit disc and any points z and w in \mathbb{S}^1 , we have*

$$(23) \quad \frac{|g(z) - g(w)|}{|z - w|} = |g^\#(z)|^{1/2} |g^\#(w)|^{1/2}$$

and

$$(24) \quad |g^\#(z)| = \frac{1 - |g(0)|^2}{|z - g^{-1}(0)|^2}.$$

The following notation will be useful in what follows. For two sequences of positive real numbers, x_n and y_n , for $n \in \mathbb{N}$, we write $x_n \asymp y_n$ if there exists positive real constants A and B such that $A \leq x_n/y_n \leq B$ for all large enough n .

LEMMA 6.4. *We have $|X_j| \asymp |Y_j|$ over all $j \in \{0, 1\}^{<\omega}$.*

PROOF. This statement depends on the fact that the backward and forward limit sets of $S = \langle f_0, f_1 \rangle$ are disjoint.

Since $|X_j| = |F_j(a) - F_j(b)|$ we obtain

$$(25) \quad |X_j| = \frac{(1 - |F_j(0)|^2)|X|}{|F_j^{-1}(0) - a||F_j^{-1}(0) - b|}$$

upon combining (23) and (24).

Similarly, we have

$$|Y_j| = \frac{(1 - |F_j(0)|^2)|Y|}{|F_j^{-1}(0) - f_1(a)||F_j^{-1}(0) - f_0(b)|}$$

and so

$$\frac{|Y_j|}{|X_j|} = \frac{|Y||F_j^{-1}(0) - a||F_j^{-1}(0) - b|}{|X||F_j^{-1}(0) - f_1(a)||F_j^{-1}(0) - f_0(b)|}.$$

Since the points $a, b, f_1(a)$ and $f_0(b)$ do not lie in the backward limit set of S , each term in the numerator and denominator of the right-hand side is bounded away from 0, uniformly over all $j \in \{0, 1\}^{<\omega}$. \square

Corollary 6.2 shows that the set of $x \in X$ that have infinitely many addresses is topologically a large set. The next theorem shows that this set is large in a measure theoretic sense too. For any $x \in \mathbb{C}$ and $r > 0$ we shall use the notation $B_x(r)$ to denote the Euclidean open ball centred at x and of radius r . We let μ denote the Lebesgue measure on \mathbb{S}^1 , normalised so that $\mu(\mathbb{S}^1) = 1$.

THEOREM 6.5. *Let f_0 and f_1 be two loxodromic elements in $\text{Aut}(\mathbb{D})$ that are not antiparallel, and let X and Y be the sets described above, which depend on f_0 and f_1 . If Y is non-empty, then $\mu(E_{\geq \aleph_0}) = \mu(X)$.*

PROOF. By Lemma 6.1 we have

$$X \setminus E_{\geq \aleph_0} = \bigcup_{n=0}^{\infty} \left(X \setminus \bigcup_{\substack{k=n \\ i \in \{0,1\}^k}}^{\infty} Y_i \right).$$

We write

$$Q_n = X \setminus \bigcup_{\substack{k=n \\ i \in \{0,1\}^k}}^{\infty} Y_i$$

and show that each Q_n has zero measure. Suppose towards contradiction that some Q_n has positive measure, then by the Lebesgue density theorem 4.7 there exists $x \in Q_n$ such that

$$(26) \quad \lim_{r \rightarrow 0} \frac{\mu(Q_n \cap B_x(r))}{\mu(B_x(r))} = 1.$$

Since x lies in X , x has at least one address and so we can choose some $i \in \{0,1\}^\omega$ such that $x = x(i)$. For each positive integer m we write $r_m = |X_{i \upharpoonright m}|$. Since $x(i) \in X_{i \upharpoonright m}$ we have $Y_{i \upharpoonright m} \subseteq X_{i \upharpoonright m} \subseteq B_x(r_m)$ for all $m \geq n$. Since $Q_n \subseteq X \setminus Y_{i \upharpoonright m}$ and $\mu(B_x(r_m)) = 2|X_{i \upharpoonright m}|$ we have

$$\frac{\mu(Q_n \cap B_x(r_m))}{\mu(B_x(r_m))} \leq \frac{2|X_{i \upharpoonright m}| - |Y_{i \upharpoonright m}|}{2|X_{i \upharpoonright m}|}$$

for all $m \geq n$. By Lemma 6.4 there is some constant $c > 0$ such that for all m we have $|Y_{i \upharpoonright m}|/|X_{i \upharpoonright m}| \geq c$. Hence

$$\frac{\mu(Q_n \cap B_x(r_m))}{\mu(B_x(r_m))} \leq 1 - \frac{c}{2}$$

for all $m \geq n$. Upon taking the limit as $m \rightarrow \infty$, we obtain a contradiction with (26), and so Q_n must have zero measure after all. \square

LEMMA 6.6. *Let f_0 and f_1 be two loxodromic elements in $\text{Aut}(\mathbb{D})$ that are not antiparallel, then we have*

$$E_{> \aleph_0} = f_0(E_{> \aleph_0}) \cup f_1(E_{> \aleph_0}).$$

PROOF. First suppose $z \in E_{> \aleph_0}$, that is z has uncountably many addresses. This means that there is an uncountable set $C \subseteq \{0,1\}^\omega$ such that for all $i \in C$ we have $z = x((i_0, \dots, i_n, \dots))$. Hence for $j = 0, 1$ we have $f_j(z) = x((j, i_0, \dots, i_n, \dots))$ for all $i \in C$. This means that $f_j(z) \in E_{> \aleph_0}$ and so $E_{> \aleph_0} \supseteq f_0(E_{> \aleph_0}) \cup f_1(E_{> \aleph_0})$. Conversely,

again suppose $z \in E_{>\aleph_0}$ so that there is an uncountable set $C \subseteq \{0, 1\}^\omega$, such that for all $i \in C$ we have $z = x((i_0, \dots, i_n, \dots))$. By the pigeon-hole principle, either $i_0 = 0$ or $i_0 = 1$ for uncountably many $i \in C$; hence either $f_0^{-1}(z) \in E_{>\aleph_0}$ or $f_1^{-1}(z) \in E_{>\aleph_0}$. In other words $z \in f_0(E_{>\aleph_0}) \cup f_1(E_{>\aleph_0})$. \square

THEOREM 6.7. *Let f_0 and f_1 be two loxodromic elements in $\text{Aut}(\mathbb{D})$ that are not antiparallel, and let X and Y be the sets described above, which depend on f_0 and f_1 . If Y is non-empty and $E_{>\aleph_0}$ is measurable, then either $\mu(E_{>\aleph_0}) = \mu(X)$ or $\mu(E_{\aleph_0}) = \mu(X)$.*

PROOF. By Lemma 6.1 we have $X = \Lambda^+$, and since the forward and backward limit sets are disjoint, $\Lambda^+ = \Lambda_c^+$ by Theorem 4.18. Furthermore, by Lemma 6.6, $E_{>\aleph_0}$ is forward invariant. Notice that $E_{\leq \aleph_0}$ is backward invariant since $E_{>\aleph_0}$ is forward invariant. Hence it follows from Lemma 4.8 (applied to S^{-1} rather than S) that $\mu(\Lambda_c^+ \cap E_{\leq \aleph_0}) > 0$ implies $\mu(E_{\leq \aleph_0}) = 1$. That is if $\mu(X \cap E_{\leq \aleph_0}) > 0$, then $\mu(X \cap E_{\leq \aleph_0}) = \mu(X)$, and so $\mu(E_{\aleph_0}) = \mu(X)$ by Theorem 6.5. Now either $\mu(X \cap E_{\aleph_0}) > 0$ so that $\mu(E_{\aleph_0}) = \mu(X)$, otherwise $\mu(X \cap E_{\aleph_0}) = 0$, and so $\mu(E_{\aleph_0}) = 0$ since $E_{\aleph_0} \subseteq X$. \square

If S is not free then we shall show that $E_{>\aleph_0}$ is not empty. When $Y = f_0(X) \cap f_1(X)$ has empty interior the semigroup $S = \langle f_0, f_1 \rangle$ is free. We show that for at least some choices of f_0 and f_1 , S is not free. We go on to show that whenever S is not free, $\mu(E_{2\aleph_0}) = \mu(X)$. We first observe that by conjugating, we can ensure that the axes of the generators are symmetrical about both the real and imaginary axes, as shown in Figure 6.1. Hence there are only three variables in \mathbb{R}^+ that determine S up to conjugation: the distance between the axes of the generators and the two translation lengths of the generators. Let ℓ_h and ℓ_v denote the geodesics contained in the horizontal and vertical axes respectively. Let σ_h and σ_v denote the reflections in ℓ_h and ℓ_v respectively. If w is the element

$$w = (i_0, \dots, i_{n-1})$$

in $\{0, 1\}^{<\omega}$ then we let w^* denote the element obtained from w by changing each instance of 0 to 1 and each instance of 1 to 0. That is

$$w^* = (1 - i_0, \dots, 1 - i_{n-1}).$$

We let w^\dagger be the element obtained from w by reversing the order of each letter of w . That is

$$w^\dagger = (i_{n-1}, \dots, i_0).$$

We say that an element w in $\{0, 1\}^{<\omega}$ is *palindromic* if $w = w^\dagger$. For a loxodromic element g , we let $\text{Ax}(g)$ denote the axis of g with its direction marked. This is in contrast with the notation $\text{ax}(g)$, which denotes the axis of g as a set.

LEMMA 6.8. *If w is an element in $\{0, 1\}^{<\omega}$, then $\sigma_h F_w \sigma_h = F_{w^\dagger}^{-1}$. Moreover, if f_0 and f_1 have the same translation lengths, then $\sigma_v F_w \sigma_v = F_{w^*}$.*

PROOF. By considering the action of $\sigma_h f_i \sigma_h$ on the point of intersection of $\text{ax}(f_i)$ and ℓ_h we see that $\sigma_h f_i \sigma_h = f_i^{-1}$ for $i = 0, 1$. Hence the first statement follows upon observing that if $g = f_{i_1} f_{i_2} \dots f_{i_{n-1}} f_{i_n}$ then

$$\sigma_h g \sigma_h = (\sigma_h f_{i_1} \sigma_h)(\sigma_h f_{i_2} \sigma_h) \dots (\sigma_h f_{i_{n-1}} \sigma_h)(\sigma_h f_{i_n} \sigma_h).$$

The second statement is similar. Noting that both generators have the same translation length and considering, for example, the action of $\sigma_v f_i \sigma_v$ on the intersection point of $\text{ax}(f_{1-i})$ and ℓ_h , we see that $\sigma_v f_i \sigma_v = f_{1-i}$. \square

Suppose now that the generators have equal translation lengths and g is an element of S such that $g = F_w$ for some palindromic element w in $\{0, 1\}^{<\omega}$, and write $g^* = F_{w^*}$. For example g could be $f_0 f_1 f_0$, and so $g^* = f_1 f_0 f_1$. A consequence of the first statement in the lemma is that $\text{Ax}(g) = \sigma_h(\text{Ax}(g^{-1}))$. It follows that $\text{Ax}(g)$ is orthogonal to ℓ_h . Similarly, by the second statement in the lemma we see that $\text{Ax}(g) = \sigma_v(\text{Ax}(g^*))$. Figure 6.2 shows an example in the case where the axes of f_0 and f_1 do not meet.

Since $\sigma_v g \sigma_v = g^*$ it follows that g and g^* have the same translation lengths. This means that if we can find g induced by a palindromic element w such that $w \neq w^*$ and whose axis is equal to ℓ_h , then $g \in S$ is induced by at least two different elements of $\{0, 1\}^{<\omega}$, and so S is not a free semigroup. Towards this, we consider varying the common translation length τ of the generators, and consider g and g^* as dependant on τ . Choose τ small enough such that $f_0(b)$ is in the right half-plane. Now choose n large enough such that $f_1^n f_0(0)$ is

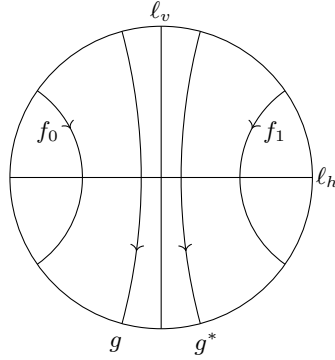


FIGURE 6.2. The axis of an element g in S induced by a palindromic element.

close enough to $f_0(b)$ (where b is the attracting fixed point of f_1) such that $f_0 f_1^n f_0(0)$ is in the right half-plane, and close to the ideal boundary. It follows that the attracting fixed point of $f_0 f_1^n f_0$ is in the right half-plane. We put $g_n = f_0 f_1^n f_0$ and conclude that the axis of g_n is contained in the right half-plane. Now, with n fixed, we see that as τ tends to ∞ the attracting fixed point of g_n converges to a , the attracting fixed point of f_0 . Hence for large τ , the axis of g is in the left half-plane. Since the attracting and repelling fixed points of g vary continuously with τ we see by the intermediate value theorem that for some τ the axis of g coincides with ℓ_v . In this case we must have $g = g^*$ as g and g^* have the same translation lengths. Since w and w^* are distinct and $g = F_w = F_{w^*}$, we see that S is not a free semigroup.

The above example of a semigroup that is not free was found by relying on symmetry; we do not know if a ‘typical’ choice of f_0 and f_1 gives rise to a semigroup that is not free. However, whenever a semigroup is not free, we shall see that the set of points in X with 2^{\aleph_0} -many addresses has the same measure as X .

For any p in $\{0, 1\}^{<\omega}$, let $A(p)$ be the set of points $x \in X$ such that every address for x eventually avoids p . That is, $A(p)$ is the set of those x in X with the property that for all $i \in \{0, 1\}^\omega$ such that $x(i) = x$, we have $(i_{n+1}, \dots, i_{n+|p|}) = p$ for only finitely many $n \in \mathbb{N}$. Recall that for i and j in $\{0, 1\}^{<\omega}$ we let i, j denote the concatenation of i and j .

LEMMA 6.9. *For every p in $\{0, 1\}^{<\omega}$, the set $A(p)$ has measure zero.*

PROOF. The proof is very similar to that of Theorem 6.5. We let $A_m(p)$ denote the set of points $x \in X$ such that every address of x avoids p after m letters. That is, if $x = x(i)$ for some $i \in \{0,1\}^\omega$, then $i \upharpoonright m$, that is (i_0, \dots, i_{m-1}) , does not contain p . Hence

$$A_m(p) = X \setminus \bigcup_{\substack{i \in \{0,1\}^m \\ j \in \{0,1\}^{<\omega}}} X_{i,j,p}.$$

We now suppose towards contradiction that $A_m(p)$ has positive measure, so that by the Lebesgue density theorem there exists a point $x \in A_m(p)$ such that

$$(27) \quad \lim_{r \rightarrow 0} \frac{\mu(A_m(p) \cap B_x(r))}{\mu(B_x(r))} = 1,$$

where as before, $B_x(r)$ is the ball of radius r centred at x . Choose some address i of x . For each positive integer n we write $r_n = |X_{i \upharpoonright n}|$. Since $x(i) \in X_{i \upharpoonright n}$ we have, for all $n \geq m$, $X_{i \upharpoonright n, p} \subseteq X_{i \upharpoonright n} \subset B_x(r_n)$. Hence we have

$$\frac{\mu(A_m(p) \cap B_x(r_n))}{\mu(B_x(r_n))} \leq \frac{2|X_{i \upharpoonright n}| - |X_{i \upharpoonright n, p}|}{2|X_{i \upharpoonright n}|}$$

for all $n \geq m$. We claim there is a constant $c > 0$ such that for all n we have $|X_{i \upharpoonright n, p}|/|X_{i \upharpoonright n}| \geq c$. Assume for now the claim is true, then

$$\frac{\mu(A_m(p) \cap B_x(r_n))}{\mu(B_x(r_n))} \leq 1 - \frac{c}{2}$$

for all $n \geq m$. Upon taking the limit as $n \rightarrow \infty$, we obtain a contradiction with (27), and so $A_m(p)$ must have zero measure after all. It follows that

$$A(p) = \bigcup_{m \in \omega} A_m(p)$$

also has zero measure.

In order to verify the constant c exists as claimed, we apply equation (25) from the proof of Lemma 6.4 to give

$$\frac{|X_{i \upharpoonright n, p}|}{|X_{i \upharpoonright n}|} \asymp \frac{1 - |F_{i \upharpoonright n, p}(0)|^2}{1 - |F_{i \upharpoonright n}(0)|^2}.$$

In the disc model, for any $x \in \mathbb{D}$ we have

$$\rho(0, x) = \log \left(\frac{1 + |x|}{1 - |x|} \right)$$

(see for example [34, Section 1.6]), and so $\exp[-\rho(0, x)] \asymp 1 - |x|$ as x runs along any sequence of points tending to the ideal boundary. Moreover, for any Möbius transformations F and g we have $\rho(Fg(0), 0) = \rho(g(0), F^{-1}(0))$ and

$$\rho(F^{-1}(0), 0) - \rho(f(0), 0) \leq \rho(g(0), F^{-1}(0)) \leq \rho(F^{-1}(0), 0) + \rho(f(0), 0).$$

It follows that as $n \rightarrow \infty$,

$$1 - |F_{i|n}(0)|^2 \asymp \exp[-\rho(F_{i|n}(0), 0)] \asymp \exp[-\rho(F_{i|n,p}(0), 0)] \asymp 1 - |F_{i|n,p}(0)|^2.$$

It follows that $|X_{i|n,p}|/|X_{i|n}|$ is bounded between two positive real numbers for all large enough n , and so the constant c exists as claimed. \square

LEMMA 6.10. *Suppose S is not free, say for some distinct p and q in $\{0, 1\}^{<\omega}$ we have $F_p = F_q$. Then, for any $i \in \{0, 1\}^\omega$ that does not eventually avoid p , the point $x(i)$ in S has 2^{\aleph_0} -many addresses; that is, $x(i) \in E_{2^{\aleph_0}}$.*

PROOF. Suppose $i \in \{0, 1\}^\omega$ takes the form $i = j_1, p, j_2, p, \dots, j_n, p, \dots$ where each $j_k \in \{0, 1\}^{<\omega}$. This representation need not be unique, and possibly some j_k have length 0. For any $k \in \{0, 1\}^\omega$ consider the element $i(k) \in \{0, 1\}^\omega$ obtained from i by replacing the instance of p immediately after j_n with q if $k(n) = 1$; otherwise we leave it as p . If the lengths $|p|$ and $|q|$ of p and q (respectively) are equal, it is clear that the map $k \mapsto i(k)$ is one-to-one. Next consider the case where these lengths are not equal, say $|p| < |q|$. Suppose towards contradiction that $k \mapsto i(k)$ is not one-to-one, say k and k' are distinct elements in $\{0, 1\}^\omega$ and yet $i(k) = i(k')$. If m is the least positive integer such that $k(m)$ disagrees with $k'(m)$, then we have

$$j_1, \dots, j_m, p, \dots = j_1, \dots, j_m, q, \dots$$

In particular $q \upharpoonright |p| = p$, but then as $F_q = F_p$ we have $I = f_{q|p|} \dots f_{q|q|-1}$, which is impossible. Hence in all cases, the map $k \mapsto i(k)$ is one-to-one. It follows that there are 2^{\aleph_0} -many addresses for $x(i)$. \square

The discussion following Lemma 6.8 shows that there exist choices of f_0 and f_1 such that S is not free. The next theorem tells us that, for such semigroups, almost every point in X has 2^{\aleph_0} -many addresses.

THEOREM 6.11. *If $S = \langle f_0, f_1 \rangle$ is not free, then $E_{2^{\aleph_0}}$ has the same measure as X .*

PROOF. As S is not free, there are distinct p and q in $\{0, 1\}^{<\omega}$ such that $F_p = F_q$. By Lemma 6.9 the set of points for which every address eventually avoids the block p has measure zero. Hence $X \setminus A(p)$, the set of points in X for which there exists some address that contains infinitely many instances of the block p , has the same measure as X . By Lemma 6.10 $X \setminus A(p)$ is contained in $E_{2^{\aleph_0}}$, and so $E_{2^{\aleph_0}}$ also has the same measure as X . \square

Although the theorem above can be generalised to higher dimensions, it is not clear that it can be generalised to cases where $\Lambda^+(S)$ has non-integer Hausdorff dimension; for if s is the non-integer Hausdorff dimension of Λ^+ , then the direct analogue of the Lebesgue density theorem fails spectacularly when Λ^+ has finite and nonzero s -measure.

Bibliography

- [1] B. Aebischer, *The limiting behavior of sequences of Möbius transformations*, Math. Z. **205** (1990), no. 1, 49–59.
- [2] I. Bárány, A. F. Beardon, and T. K. Carne, *Barycentric subdivision of triangles and semigroups of Möbius maps*, Mathematika **43** (1996), no. 1, 165–171.
- [3] A. F. Beardon, *The geometry of discrete groups*, Graduate Texts in Mathematics, vol. 91, Springer-Verlag, New York, 1995. Corrected reprint of the 1983 original.
- [4] ———, *Continued fractions, discrete groups and complex dynamics*, Comput. Methods Funct. Theory **1** (2001), no. 2, 535–594.
- [5] ———, *The Hillam-Thron theorem in higher dimensions*, Geom. Dedicata **96** (2003), 205–209.
- [6] A. F. Beardon, M. Hockman, and I. Short, *Geodesic continued fractions*, Michigan Math. J. **61** (2012), 133–150.
- [7] A. F. Beardon and L. Lorentzen, *Continued fractions and restrained sequences of Möbius maps*, Rocky Mountain J. Math. **34** (2004), no. 2, 441–466.
- [8] P. Billingsley, *Ergodic theory and information*, John Wiley & Sons, Inc., New York-London-Sydney, 1965.
- [9] B. Branner and N. Fagella, *Quasiconformal Surgery in Holomorphic Dynamics*, Cambridge studies in advanced mathematics, vol. 141, Cambridge University Press, Cambridge, 2014.
- [10] C. R. Bell, *Non-discrete Möbius groups*, Master’s thesis, University of Cambridge, 1997.
- [11] E. Crane and I. Short, *Conical limit sets and continued fractions*, Conform. Geom. Dyn. **11** (2007), 224–249 (electronic).
- [12] T. Das, D. Simmons, and M. Urbański, *Geometry and dynamics in Gromov hyperbolic metric spaces*, <http://arxiv.org/abs/1409.2155>.
- [13] K. Falconer, *Techniques in fractal geometry*, John Wiley & Sons, Ltd., Chichester, 1997.
- [14] K. J. Falconer, *The geometry of fractal sets*, Cambridge Tracts in Mathematics, vol. 85, Cambridge University Press, Cambridge, 1986.
- [15] D. Fried, S. M. Marotta, and R. Stankewitz, *Complex dynamics of Möbius semigroups*, Ergodic Theory Dynam. Systems **32** (2012), no. 6, 1889–1929.
- [16] S. Friedland, *Computing the Hausdorff dimension of subshifts using matrices*, Linear Algebra Appl. **273** (1998), 133–167.

- [17] J. Gilman, *Two-generator discrete subgroups of $\mathrm{PSL}(2, \mathbf{R})$* , Mem. Amer. Math. Soc. **117** (1995), no. 561, x+204.
- [18] J. Gilman and B. Maskit, *An algorithm for 2-generator Fuchsian groups*, Michigan Math. J. **38** (1991), no. 1, 13–32.
- [19] T. Harada, *The dynamics of nearly abelian polynomial semigroups at infinity*, Proc. Japan Acad. Ser. A Math. Sci. **73** (1997), no. 3, 36–37.
- [20] A. Hinkkanen and G. J. Martin, *The dynamics of semigroups of rational functions. I*, Proc. London Math. Soc. (3) **73** (1996), no. 2, 358–384.
- [21] M. Jacques and I. Short, *Continued fractions and semigroups of Möbius transformations*, <http://arxiv.org/abs/1609.00576>.
- [22] ———, *Repeated compositions of Möbius transformations*, <http://arxiv.org/abs/1609.01261>.
- [23] W. Just and M. Weese, *Discovering modern set theory. II. Set-theoretic tools for every mathematician*, Graduate Studies in Mathematics, vol. 18, American Mathematical Society, Providence, RI, 1997.
- [24] H. Kriete and H. Sumi, *Semihyperbolic transcendental semigroups*, J. Math. Kyoto Univ. **40** (2000), no. 2, 205–216.
- [25] J. Lehner, *Discontinuous groups and automorphic functions*, Mathematical Surveys, No. VIII, American Mathematical Society, Providence, R.I., 1964.
- [26] L. Lorentzen, *General convergence of continued fractions*, Trans. Amer. Math. Soc. **294** (1986), no. 2, 477–485.
- [27] ———, *Möbius transformations mapping the unit disk into itself*, Ramanujan J. **13** (2007), no. 1-3, 253–263.
- [28] L. Lorentzen and H. Waadeland, *Continued fractions. Vol. 1. Convergence theory*, 2nd ed., Atlantis Studies in Mathematics for Engineering and Science, vol. 1, Atlantis Press, Paris; World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2008.
- [29] A. Marden, *Outer circles. An introduction to hyperbolic 3-manifolds*, Cambridge University Press, Cambridge, 2007.
- [30] B. Maskit, *On Klein’s combination theorem*, Trans. Amer. Math. Soc. **120** (1965), 499–509.
- [31] J. P. Matelski, *The classification of discrete 2-generator subgroups of $\mathrm{PSL}(2, \mathbf{R})$* , Israel J. Math. **42** (1982), no. 4, 309–317.
- [32] J. Milnor, *Dynamics in one complex variable*, 3rd ed., Annals of Mathematics Studies, vol. 160, Princeton University Press, Princeton, NJ, 2006.
- [33] D. Mumford, C. Series, and D. Wright, *Indra’s pearls. The vision of Felix Klein*, Cambridge University Press, New York, 2002.
- [34] P. J. Nicholls, *The ergodic theory of discrete groups*, London Mathematical Society Lecture Note Series, vol. 143, Cambridge University Press, Cambridge, 1989.

- [35] I. Short, *Hausdorff dimension of sets of divergence arising from continued fractions*, Proc. Amer. Math. Soc. **140** (2012), no. 4, 1371–1385.
- [36] R. Stankewitz, *Completely invariant Julia sets of polynomial semigroups*, Proc. Amer. Math. Soc. **127** (1999), no. 10, 2889–2898.
- [37] ———, *Uniformly perfect sets, rational semigroups, Kleinian groups and IFS's*, Proc. Amer. Math. Soc. **128** (2000), no. 9, 2569–2575.
- [38] H. Sumi, *On dynamics of hyperbolic rational semigroups*, J. Math. Kyoto Univ. **37** (1997), no. 4, 717–733.
- [39] Y. Sun and C.-C. Yang, *On the connectivity of the Julia set of a finitely-generated rational semigroup*, Proc. Amer. Math. Soc. **130** (2002), no. 1, 49–52 (electronic).